

# Exploiting Explicit Annotations and Semantic Types for Implicit Argument Resolution

Egoitz Laparra  
IXA Group  
Basque Country University  
San Sebastian, Spain  
Email: egoitz.laparra@ehu.es

German Rigau  
IXA Group  
Basque Country University  
San Sebastian, Spain  
Email: german.rigau@ehu.es

**Abstract**—Following the frame semantics paradigm, we present a novel strategy for solving null-instantiated arguments. Our method learns probability distributions of semantic types for each Frame Element from explicit corpus annotations. These distributions are used to select the most probable missing implicit arguments together with its most probable fillers. We empirically demonstrate that our method outperforms the systems evaluated on the SemEval 2010 task 10 dataset.

## I. INTRODUCTION

One of the most relevant part of semantic processing of text is identifying the overtly realized arguments of a predicate. Although several systems have been developed to perform this task, called Semantic Role Labeling (SRL) [1], they have traditionally focused on searching the fillers for the arguments in the local context of the predicate, that is, within those participants that share a syntactical relation with the predicate. Since traditional SRL systems depend strongly on those syntactic relations, they cannot perform annotations when the candidate instantiation of the argument is not explicit. However, null instantiated arguments can be inferred from the context and the explicit realization of the other arguments. Using the nominal predicates of NomBank [2], Gerber and Chai [3] pointed out that the implicit arguments can add up to 65% to the coverage of the instantiations. As a consequence, the increase of connections within the predicates and their participants could help dramatically text understanding.

In FrameNet [4], the predicates, called lexical-units (LU), evoke frames which roughly correspond to different events or scenarios. For each frame a set of possible arguments are defined. These arguments are called *Frame Elements* (FE). The *Core Frame Elements* are the essential FEs of a frame and can define it by themselves. However, not every *Core FEs* is always present in a sentence. These FEs are considered as *Null Instantiations* (NI). FrameNet classifies the NIs in three different sets depending on the licensor of the omission and the interpretation they can receive:

- (i) If the FEs are grammatically omitted, NIs are called *Constructional Null Instantiations* (CNI).
- (ii) If the omissions are licensed lexically and the fillers are inaccessible, the NIs are called *Indefinite Null Instantiations* (INI).

- (iii) When the omissions are licensed lexically and the fillers are recoverable, the NIs are called *Definite Null Instantiations* (DNI).

Thus, the task of annotating implicit arguments following the FrameNet structures focuses just on identifying and filling DNIs. In the next example, the LU **tenant**<sub>n</sub> evoking the frame **Residence** has an instantiated FE, *Resident*, whose filler is [the tenants]. The correct filler for the DNI corresponding to FE *Location*, [the house], appears two sentences before:

“Now, Mr. Holmes, with your permission, I will show you round the house.” The various bedrooms and sitting-rooms had yielded nothing to a careful search. Apparently [the **tenants**<sub>Residence</sub>]*Resident* had brought little or nothing with them. DNI<sub>Location</sub>

The main contribution of this paper is twofold. Firstly, a novel method for capturing the missing FEs that are DNIs. Secondly, a simple model to fill those DNIs that relies on the intuition that the heads of the filler of a FE tend to belong to the same semantic types. We also believe that those semantic types can be learned from the heads of the fillers of the explicit annotations allowing to have a broader training set that the one available for implicit annotations. For instance, as [hotel], [dwelling] and [house] share the same semantic type **Building** we could exploit annotations like the next ones to solve the case above:

[Single older people]<sub>Resident</sub> **occupy**<sub>Residence</sub> [unfit dwellings]<sub>Location</sub>.  
It's thought the fire may have been started by [squatters<sub>Residence</sub>]<sub>Resident</sub> [in the hotel]<sub>Location</sub>.

This paper is organized as follows. After this short introduction, the next section III we present the framework of our experiments using the model described in Section IV. Those experiments, and their corresponding results, are presented in Section V. Finally, we conclude in Section VI discussing our present and future work.

## II. RELATED WORK

Task 10 of SemEval-2010 focused on the evaluation of SRL systems based on the FrameNet paradigm<sup>1</sup> [5]. This task was divided in two different sub-tasks:

- (i) Argument annotation in a traditional SRL manner.
- (ii) Filling null instantiations over the document.

In the latter, the systems presented in the latter sutask identified those missing *Frame Elements* that were really *Null Instantiations*, classify those NI within either definite or indefinite, and finally look for the correct fillers of the DNIs. Two systems participated in the second sub-task: VENSES++ and SEMAFOR.

VENSES++ [6] builds logical rules from syntactic parsing and uses hand-crafted lexicons. For verbal predicates, NIs are resolved using a semantic similarity measure between a NI and the potential fillers using WordNet [7]. For nominal predicates, the system employs a common sense reasoning module that builds upon ConceptNet [8]. The same authors improved the performance of their system in a later work [9].

SEMAFOR [10] is a supervised system that extends an existing semantic role labeler [11] replacing the features defined for regular arguments with two new semantic features. First, it checks if a potential filler in the context fills the null-instantiated role overtly in one of the FrameNet sentences, and secondly, it calculates the distributional semantic similarity between the fillers and the roles. Although this system obtained the best performance in the task the sparseness of the data affects strongly its results.

In a different approach, Ruppenhofer et al. [12] explore some different linguistic strategies in order to enhance the DNI identification. They conclude that a more sophisticated approach for DNI identification can improve significantly the performance of the whole pipeline, even if the method for the DNI filling is simple.

More recently, Silberer and Frank [13] present a first attempt to solve the task adapting an entity-based coreference resolution model. In this work, the authors also extend automatically the training corpus to avoid the problem of data sparsity.

In another closely related work, Gerber and Chai [3] define a set of syntactic and semantic features to characterize implicit arguments for those predicates appearing in NomBank [2] to train a logistic regression classifier. Unlike Semeval 2010, the resulting dataset contains a much higher number of annotations, but just for only ten different predicates.

## III. DATA

We have used for our experiments the dataset distributed for the Task 10 in Semeval-2010, specially the NI resolution subtask data that includes a manual annotation of the explicit FEs. The corpus contains some chapters extracted from two Arthur Conan Doyle’s stories. “The Tiger of San Pedro” chapter from “The Adventure of Wisteria Lodge” was selected for training, while chapters 13 and 14 from “The Hound of the Baskervilles” were selected for testing. The texts are annotated

using the frame-semantic structure of FrameNet 1.3 including null instantiations, the type of the NI and the corresponding fillers for each DNI. As it can be seen in table I the number of null instantiations in the training set is not only small but even smaller than the number of annotations in the testing set. The same table also shows the number of explicit FE annotated in the corpus that are the ones we use to train our model.

data-set	DNIs (solved)	INIs	Explicit FE
train	303 (245)	277	2726
test-13	158 (121)	116	1545
test-14	191 (138)	245	1688

TABLE I  
NUMBER OF NI AND EXPLICIT FE ANNOTATIONS FOR THE SEMEVAL-10  
TASK-10 CORPUS

The dataset also includes the annotation files for the lexical units and the full-text annotated corpus from FrameNet.

## IV. MODEL

Chen et al. [10] complain about the sparseness of the training set. In the previous section, table I shows that the number of annotations for DNIs is low with respect to the size of the training set for explicit FEs.

We propose that the explicit annotation can be exploited for the NI resolution. In particular, this preliminary version of our model learns the semantic knowledge associated with the heads of the participants that fill the FEs. This knowledge is used to try to capture the heads of the participants that should fill the NIs. We leave for a further development of the model the definition of the correct spans of the fillers.

First of all we perform a syntactic and semantic analysis of the dataset, both training and testing parts. We use the Stanford parser<sup>2</sup> to obtain the Name Entities and coreference chains in order to process all the occurrences of the same participant as an unique item. We also perform a very simple Word Sense Disambiguation (WSD) process assigning to each word, when possible, the most frequent sense of WordNet [7]. This heuristic has been used frequently as a baseline in the evaluation of WSD systems and it seems to be very hard to beat [14]. As the senses of WordNet have been mapped to several ontologies, this disambiguation allows us to label the documents with ontological features that can work as semantic types. In this work we have used the Top Ontology (TO) [15]. For those cases where the words cannot be labelled with any feature we define the pair lemma#part-of-speech as their semantic type (see the last example in Table II). Our model assigns to each instantiated FE the ontological feature of the syntactic head of its filler. Then, our model learns from the training data the probability distribution  $\mathbf{s}$  of the semantic types of each FE. It also calculates the probability distribution  $\mathbf{p}$  of the part-of-speech of the head of their fillers. The model includes in the training data the explicit annotations from the test document which is being analyzed. Table II contains three

<sup>1</sup>[http://www.coli.uni-saarland.de/projects/semeval2010\\_FG/](http://www.coli.uni-saarland.de/projects/semeval2010_FG/)

<sup>2</sup><http://nlp.stanford.edu/software/dcoref.shtml>

Frame#FrameElement	Head	SemanticType
Expectation#Cognizer	Holmes [person]	Function#Human#Living#Object
Residence#Location	hotel	Artifact#Building#Object
Opinion#Opinion	that	that#IN

TABLE II

SOME EXAMPLES OF SEMANTIC TYPES ASSIGNED TO FES. IN BRACKETS, THE LABEL OF THE NER

different examples of this assignment, the first one corresponds to a case where the TO feature has been assigned through the NER label.

Before performing the implicit argument resolution in a document we have to decide which not instantiated *Frame Elements* should be selected. That is, which are *Definite Null Instantiations*. Since our approach relies on learning the explicit annotations, our strategy differs notably from those presented in previous works. Following with the example above, suppose we are processing the lexical unit  $tenant_n$  belonging to the frame **Residence** and the instantiated FE *Resident*:

Apparently [the **tenants<sub>Residence</sub>**]*Resident* had brought little or nothing with them. *DNI<sub>Location</sub>*

Our system collects from the training data the most common FE patterns of the corresponding frame of the LU under study. The patterns collected must contain the instantiated FEs of the LU. Table III shows the patterns collected for the previous example.

Resident Location	384
Resident Co_resident Location	34
Resident Co_resident	14
Resident	13
Resident Location Manner	1
Resident Location Time	1

TABLE III

MOST COMMON PATTERNS FOR THE FRAME **Residence** CONTAINING THE FE *Resident* ORDERED BY FREQUENCY

Then, the system defines as DNIs all the *Core Frame Elements* of the most common pattern that are missing for the lexical unit that it is being processed. In the example above the most common pattern having the FE *Resident* is the one formed by the sequence of FEs *Resident* and *Location*. As the FE *Location* is indeed a *Core FE* of the frame **Residence**, it will be defined as a DNI of  $tenant_n$  and our system will try to find a filler for it.

Once the previous process has been applied for all the lexical units in the document, our system can perform the DNI resolution. Gerber and Chai [3] showed that the vast majority of the fillers of the implicit arguments can be found within the same sentence containing the predicate or in the two previous ones. They establish a window formed by these three sentences and considers as possible candidates the participants belonging to that window. We use the same criteria in our model. Thus,

Head	PoS	SemanticType	P(s)	P(p)	P(s,p)
house	N	Artifact#Building#Object	0.164	0.376	0.062
bedroom	N	Artifact#Building#Object#Part#Place	0.005	0.376	0.002
sitting-room	N	Artifact#Building#Object#Part#Place	0.005	0.376	0.002
Holmes [person]	N	Function#Human#Living#Object	0	0.376	0
show	V	Communication#Dynamic#Experience	0	0	0

TABLE IV

RESULTING PROBABILITIES FOR SOME OF THE CANDIDATES IN THE CONTEXT OF  $tenant_n$

our system selects the filler among the terminals that belong to the three sentences, the closest one that maximizes  $P(s, p)$ , the joint probability of **s** and **p**. Following with the example, our system calculates  $P(s, p)$  for all terminals in the three sentence window of  $tenant_n$ :

“Now, Mr. Holmes, with your permission, I will show you round the house.” The various bedrooms and sitting-rooms had yielded nothing to a careful search. Apparently the tenants had brought little or nothing with them.

Table IV shows that, in this case, [house] obtains the higher joint probability. Consequently, our model selects this participant as the filler for the NI *Location* for the predicate  $tenant_n$ .

## V. EVALUATION

As explained in section IV, the first step of our model consists of the correct identification of those missing FEs that are actually DNIs. Given that the final output of the system depends strongly on this first step, we evaluated the performance of our methodology in the DNI identification process. Table V shows how our system outperforms state of the art systems on this subtask.<sup>3</sup>

System	P	R	F1
Tonelli and Delmonte, 2010 [6]	-	-	-
Chen et al., 2010 [10]	<b>0.57</b>	0.03	0.06
Tonelli and Delmonte, 2011 [9]	0.39	0.43	0.41
This work	0.50	<b>0.66</b>	<b>0.57</b>

TABLE V

EVALUATION OF DNI IDENTIFICATION.

For the second subtask, we have used the scorer provided for NI subtask for the evaluation of the DNI resolution. This scorer works slightly different that the one for the traditional SRL subtask. Since the participants can appear repeatedly over the same document, the scorer has to take into account coreference chains of the possible fillers. Thus, the gold-standard files include the full coreference annotations and the scorer counts as correct any correct mention of the filler. In this subtask, precision is defined as the number of all true positive instantiations divided by the number of solutions provided by a

<sup>3</sup>Values for the first version of VENSES++ were not reported. Silberer and Frank, [13] obtain a Recall of 0.4 in NI classification but they do not report results separately for DNI.

system, and recall as the number of true positive instantiations divided by the number of solutions in the gold standard. F-Score is the usual harmonic mean of recall and precision. Table VI presents the results of the different systems.<sup>4</sup> It includes the performance of our system when learning either the lemmas or the semantic types of the head of the fillers. These results show that there is an additional gain exploiting the semantics of the fillers.

System	P	R	F1
Tonelli and Delmonte, 2010 [6]	-	-	0.01
Chen et al., 2010 [10]	<b>0.25</b>	0.01	0.02
Tonelli and Delmonte, 2011 [9]	0.13	0.06	0.08
Silberer and Frank, 2012 [13]	0.09	0.11	0.10
This work (lemmas+pos)	0.13	0.23	0.17
This work (semantic-types+pos)	0.15	<b>0.25</b>	<b>0.19</b>

TABLE VI  
PERFORMANCE OF OUR SYSTEM COMPARED WITH THE SYSTEMS USING SEMEVAL 2010 DATASET.

Although our system obtains better results than alternative approaches, such a low figures clearly reflect its inherent difficulty. Our system clearly outperforms VENSES++ in terms of both precision and recall. SEMAFOR seems to solve much accurately a very limited number of cases. Finally, we also include the best results from [13] obtained when using for training a larger corpus extended heuristically. In fact, their results are much lower when using only the training corpus provided for the task. It is worth mentioning that a window of three sentences around the predicate sets the upperbound recall to 76% for DNIs appearing in the test documents. Extending the search for fillers beyond this window is part of a future work.

## VI. CONCLUSIONS AND FUTURE WORK

We have presented the initial results of a novel strategy for solving null-instantiated arguments. Our proposal for DNI identification present a clear improvement that positively affects the overall performance. They results also show the benefit of learning and exploiting explicitly instantiated arguments for solving null-instantiations. This seems to be a promising approximation to the problem in order to overcome the limitation of too small training datasets for this task. The results also show an small improvement when representing the argument fillers as semantic types. Interestingly, our model already improves by a large margin the best state of the art results on this dataset.

Our work has been focused on searching the heads of the proper fillers. In the near future, we should explore strategies to also find the complete span of those fillers. We also plan to study the performance of more sophisticated Word Sense Disambiguation systems and different knowledge resources as semantic types. Finally, we also plan to test the system on the NomBank dataset [3].

<sup>4</sup>The values of P and R for the first version of VENSES++ were not provided.

## VII. ACKNOWLEDGMENTS

This work has been supported by KNOW-2 (TIN2009-14715-C04-01). We want to thank the anonymous reviewers for their valuable comments.

## REFERENCES

- [1] D. Gildea and D. Jurafsky, "Automatic labeling of semantic roles," in *ACL*, 2000.
- [2] A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman, "The nombank project: An interim report," in *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, A. Meyers, Ed. Boston, Massachusetts, USA: Association for Computational Linguistics, May 2 - May 7 2004, pp. 24–31.
- [3] M. Gerber and J. Y. Chai, "Beyond nombank: a study of implicit arguments for nominal predicates," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ser. ACL '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 1583–1592.
- [4] C. F. Baker, C. J. Fillmore, and J. B. Lowe, "The berkeley framenet project," in *COLING-ACL*, 1998, pp. 86–90.
- [5] J. Ruppenhofer, C. Sporleder, R. Morante, C. Baker, and M. Palmer, "Semeval-2010 task 10: linking events and their participants in discourse," in *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, ser. DEW '09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 106–111.
- [6] S. Tonelli and R. Delmonte, "Venses++: Adapting a deep semantic processing system to the identification of null instantiations," in *Proceedings of the 5th International Workshop on Semantic Evaluation*, ser. SemEval '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 296–299.
- [7] C. Fellbaum, Ed., *WordNet: an electronic lexical database*. MIT Press, 1998.
- [8] H. Liu and P. Singh, "Conceptnet: A practical commonsense reasoning toolkit," *BT Technology Journal*, vol. 22, no. 4, pp. 211–226, 2004.
- [9] S. Tonelli and R. Delmonte, "Desperately seeking implicit arguments in text," in *Proceedings of the ACL 2011 Workshop on Relational Models of Semantics*. Portland, Oregon, USA: Association for Computational Linguistics, June 2011, pp. 54–62.
- [10] D. Chen, N. Schneider, D. Das, and N. A. Smith, "Semafor: Frame argument resolution with log-linear models," in *Proceedings of the 5th International Workshop on Semantic Evaluation*, ser. SemEval '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 264–267.
- [11] D. Das and N. A. Smith, "Semi-supervised frame-semantic parsing for unknown predicates," in *ACL*, 2011, pp. 1435–1444.
- [12] J. Ruppenhofer, P. Gorinski, and C. Sporleder, "In search of missing arguments: A linguistic approach," in *Recent Advances in Natural Language Processing, RANLP 2011, 12-14 September, 2011, Hissar, Bulgaria*, G. Angelova, K. Bontcheva, R. Mitkov, and N. Nicolov, Eds. RANLP 2011 Organising Committee, 2011, pp. 331–338.
- [13] C. Silberer and A. Frank, "Casting implicit role linking as an anaphora resolution task," in *\*SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*. Montréal, Canada: Association for Computational Linguistics, 7-8 June 2012, pp. 1–10.
- [14] W. A. Gale, K. W. Church, and D. Yarowsky, "Estimating upper and lower bounds on the performance of word sense disambiguation," in *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*. ACL, 1992.
- [15] J. Álvarez, J. Atserias, J. Carrera, S. Climent, E. Laparra, A. Oliver, and G. Rigau, "Complete and consistent annotation of wordnet using the top concept ontology," in *LREC*, 2008.