

# In Search of Missing Arguments: A Linguistic Approach

Josef Ruppenhofer

Dept. of Information Science  
and Language Technology  
Hildesheim University  
ruppenho@uni-hildesheim.de

Philip Gorinski and Caroline Sporleder

Computational Linguistics  
Saarland University  
{philipg,csporled}@coli.uni-saarland.de

## Abstract

Semantic argument structures are often incomplete in that core arguments are not locally instantiated. However, many of these implicit arguments can be linked to referents in the wider context. In this paper we explore a number of linguistically motivated strategies for identifying and resolving such null instantiations (NIs). We show that a more sophisticated model for identifying definite NIs can lead to noticeable performance gains over the state-of-the-art for NI resolution.

## 1 Introduction

Semantic Role Labeling (SRL) is traditionally concerned with identifying the overtly realized arguments of a predicate. However, in a natural discourse only a relatively small proportion of the theoretically possible semantic arguments tend to be locally instantiated in the same clause or sentence that contains the target predicate. The other arguments are so-called *null instantiations* (NIs). Even core arguments of a predicate, i.e., those that express participants which are necessarily present in the situation which the predicate evokes (see Section 2 for a more detailed explanation of core vs. peripheral arguments), are frequently not instantiated in the local context. While null instantiated arguments are not locally realized, they can often be inferred from the context.

Consider examples (1) and (2) below (taken from Arthur Conan Doyle’s “The Adventure of Wisteria Lodge” and part of the SemEval-10 Task-10 corpus (Ruppenhofer et al., 2010)). We use A and B in the examples to indicate speakers.<sup>1</sup> In a frame-semantic analysis of (1) *interesting* evokes the `Mental_stimulus_stimulus_focus`

<sup>1</sup>We provide this information for clarity, it is not explicitly marked in the corpus.

(`MSSF`) frame. This frame has two core semantic arguments, `EXPERIENCER` and `STIMULUS`, as well as eight peripheral arguments, such as `TIME`, `MANNER`, `DEGREE`. Of the two core arguments, neither is actually realized in the same sentence. Only the peripheral argument `DEGREE` (`DEG`) is instantiated and realized by *most*. To fully understand the sentence, it is necessary to infer the fillers of the `EXPERIENCER` and `STIMULUS` roles, i.e., the reader needs to make an assumption about what is interesting and to whom. For humans this inference is easy to make as the `EXPERIENCER` (`EXP`) and `STIMULUS` (`STIM`) roles are actually filled by *he* and *a white cock* in the previous sentence. (Note that the two utterances in (1) are spoken by the same person.) Similarly, in (2) *right* evokes the `Correctness` (`CORR`) frame, which has four core arguments, only one of which is filled locally, namely `SOURCE` (`SRC`), which is realized by *You* (and co-referent with *Mr. Holmes*). However, another argument, `INFORMATION` (`INF`), is filled by the preceding sentence (spoken by a different speaker, namely Holmes), which provides details of the fact about which Holmes was right.

- (1) A. [“A white cock,”]<sub>Stim</sub> said [he]<sub>Exp</sub>. “[Most]<sub>Deg</sub> interesting<sub>Mssf</sub>!”
- (2) A. [“Your powers seem superior to your opportunities.”]<sub>Inf</sub>  
B. “[You]<sub>Src</sub>’re right<sub>Corr</sub>, Mr. Holmes.”

While humans have no problem inferring uninstantiated roles that can be filled from the linguistic context, this is beyond the capacity of state-of-the-art semantic role labeling systems, which tacitly ignore all roles that are not instantiated locally. SRL systems thus disregard much argument-level information that is potentially necessary for solving text understanding tasks such as question answering or information extraction. That the problem of locally unrealized roles is not restricted

to the genre of narrative texts as in the examples above is evidenced by a study by Gerber and Chai (2010) who annotated implicit roles for a set of high frequency nouns in NomBank, which provides predicate argument structure annotation for nominals in the Wall Street Journal portion of the Penn Treebank. They found that implicit arguments add another 65% to the coverage of overtly instantiated roles in NomBank. Hence, the problem also arises in the news domain, at least with nominal arguments, which tend to realize fewer roles overtly due to a more restrictive syntax.

Intuitively, it is not surprising that even core arguments often remain locally unexpressed since a coherent discourse is not a collection of sentences expressing random states-of-affairs but typically is concerned with a limited set of situations which tend to be interconnected. Hence, it is unlikely that an evocation of a situation in a given sentence immediately provides exhaustive information about all possible participants. It is much more likely that this information is spread out over several sentences. Traditional, sentence- or clause-based SRL is therefore clearly a simplification, albeit one that is useful as a first approximation.

In this paper, we propose a number of strategies for identifying implicit arguments and inferring their antecedents from the context. Our aim is not so much to provide a perfect system that gives the best possible performance; rather our work is of an exploratory nature. We investigate different linguistically motivated strategies for dealing with null instantiated arguments and thereby hope to shed more light on the nature of such arguments as well as evaluating potential avenues for future research on automatically inferring referents for such arguments.

This paper is structured as follows. In the next section we provide an overview of how FrameNet models semantic argument structures and null instantiations. Section 3 discusses previous approaches to null instantiation resolution. In Section 4 we describe the data we used in our experiments. The following two sections (5 and 6) describe our model and the experiments. Finally, we conclude in 7.

## 2 Arguments and Null Instantiations in FrameNet

A predicate argument structure in FrameNet consists of a *frame* evoked by a target predicate.

Each frame defines a number of potentially possibly arguments or *frame elements* (FEs). For some FEs, FrameNet explicitly specifies a *semantic type*. For instance, the EXPERIENCER of the Mental\_stimulus\_stimulus\_focus frame (see ex. 1) is defined to be of type ‘sentient’. We make use of this information in the experiments. The set of FEs is split into core arguments, peripheral arguments, and extra-thematic arguments. *Core arguments* are seen as essential components of a frame; they distinguish the frame from other frames and represent participants which are necessarily present the situation evoked by the frame, though they may not be overtly realized in a given context. *Peripheral arguments* are optional and generalize across frames, in that they can be found in all semantically appropriate frames. Typical examples are TIME or MANNER. Finally, *extra-thematic arguments* are those that situate the event described by the target predicate against another state-of-affairs. For example, *twice* can express the extra-thematic argument ITERATION. Since only core arguments are essential to a frame, only they are analyzed as null instantiated if missing. Peripheral and extra-thematic arguments are, by definition, optional anyway.

Matters are complicated by the fact that not all core arguments of all frames can be realized simultaneously. Some frames have core arguments that are mutually exclusive. For example, in the Similarity (Sim) frame the entities being compared for similarity can either be expressed by different FEs as in (3) or collectively as in (4). The frame therefore provides three FEs ENTITY\_1 (ENT1), ENTITY\_2 (ENT2), and ENTITIES (ENTS), where the first two FEs are mutually exclusive with the third. These two sets are said to form an *exclusion set*. At the same time, ENTITY\_1 and ENTITY\_2 are said to be in a *Requires* relation, which means that occurrence of one of these two core FEs requires that the other core FE occur as well.

(3) [How]<sub>Dimension</sub> is [it]<sub>Ent1</sub> **similar**<sub>Sim</sub> [to my solution]<sub>Ent2?</sub>

(4) [They]<sub>Ents</sub> are [very]<sub>Degree</sub> **similar**<sub>Sim</sub>.

*CoreSets* define another type of relation that is important in the context of null instantiations. The idea behind CoreSets is that FEs can be interdependent, i.e., express similar semantic content, which makes it unlikely that all of them will be overtly realized in a given context. An example

are the SOURCE (SRC), PATH (PTH), and GOAL (GOAL) FEs of the MOTION (Mtn) frame. They can be expressed together as in (5) (Ruppenhofer et al., 2006) but it is more likely that only one or two of them will be expressed (6). FEs that are interdependent in such way are grouped together in CoreSets. As long as one FE from a CoreSet is expressed, none of the others is annotated as omitted. If none is expressed, the contextually most relevant one is annotated as null-instantiated.

(5) [Fred]<sub>Theme</sub> **went**<sub>Mtn</sub> [from Berkeley]<sub>Src</sub> [across North America and the Atlantic Ocean]<sub>PTH</sub> [to Paris]<sub>Goal</sub>.

(6) [Fred]<sub>Theme</sub> **went**<sub>Mtn</sub> [to Paris]<sub>Goal</sub>.

The annotation of null instantiations in SemEval-10 Task-10 follows the practice adopted by FrameNet, which is rooted in the work of Fillmore (1986). Omissions of core arguments of predicates are categorized along two dimensions, the licenser and the interpretation they receive. An NI can either be licensed by a particular lexical item or a particular grammatical construction. For example, in (7) the omission of the AUTHORITIES making the arrest is licensed by the passive construction. Such an omission can apply to any predicate with an appropriate semantics that allows it to combine with the passive construction. On the other hand, the omission in (8) is lexically specific: the verb *arrive* allows the GOAL to be unspecified but the verb *reach*, also a member of the Arriving frame, does not (9).

(7) [A drunk burglar]<sub>Spect</sub> was **arrested**<sub>Arrest</sub> after accidentally handing his ID to his victim.

(8) [We]<sub>Thm</sub> **arrived**<sub>Arrive</sub> [at 8pm]<sub>Tm</sub>.

(9) \*[We]<sub>Thm</sub> **reached**<sub>Arrive</sub> [at 8pm]<sub>Tm</sub>

The above two examples also illustrate the second major dimension of variation. Whereas, in (7) the protagonist making the arrest is only existentially bound within the discourse (an instance of indefinite null instantiation, INI), the GOAL location in (8) is an entity that must be accessible to speaker and hearer from the discourse or its context (definite null instantiation, DNI). Finally, note that the licensing construction or lexical item fully and reliably determines the interpretation. Whereas missing by-phrases have always an indefinite interpretation, whenever *arrive* omits the GOAL lexically, the GOAL has to be interpreted as definite.

As INIs do not need to be accessible within a context, the task of resolving NIs is restricted to DNIs. The complete task can then be modeled as a pipeline consisting of three sub-tasks: (i) identifying potential NIs by taking into account information about core arguments and relations between them, (ii) automatically distinguishing between DNIs and INIs by identifying NI licensing constructions or lexical items, and (iii) resolving NIs classified as DNIs to a suitable referent.

### 3 Related Work

The most closely related piece of work is the system building performed in the context of the SemEval-10 Task-10 (Ruppenhofer et al., 2010). The two participating systems which addressed the NI resolution task took very different approaches. Tonelli and Delmonte (2010) developed a knowledge-based system called VENSES++ that builds on an existing text understanding system (Delmonte, 2008). VENSES++ employs deep syntactic parsing and uses hand-crafted lexicons to generate logical forms. It then makes use of a rule-based anaphora resolution procedure before employing two different strategies for identifying and resolving NIs. For verbal predicates, argument pattern templates generated from FrameNet data are used to identify missing predicates and classify lexically licensed NIs as DNI or INI. The only type of constructionally licensed NIs that can be detected by the system are those of agents in passive constructions. NIs are resolved by reasoning about the semantic similarity between an NI and a potential filler using WordNet. For nominal predicates, the system employs a common sense reasoning module that builds upon ConceptNet (Liu and Singh, 2004). The system is conservative and has a relatively high precision, e.g., 64.2% for the DNI v. INI distinction, but a low recall, identifying less than 20% of the NIs correctly.

The second system (Chen et al., 2010) is statistical and extends an existing semantic role labeler (Das et al., 2010). The system first classifies NIs as DNI or INI and then tries to find fillers for the former. Resolving DNIs is modeled in the same way as labeling overt arguments, however the search space is extended to pronouns, NPs, and nouns outside the sentence.<sup>2</sup> When evaluating a potential filler, the syntactic features which

<sup>2</sup>This disregards other role fillers such as whole sentences as in example (2) above.

are used in argument labeling of overt arguments are replaced by two semantic features: The system checks first whether a potential filler in the context fills the null-instantiated role overtly in one of the FrameNet sentences, i.e. whether there is a precedent for a given filler-role combination among the overt arguments of the frame in FrameNet. If not, the system calculates the distributional similarity between filler and role. The surface distance between a potential filler and an NI is also taken into account. While Chen et al.’s system has a higher recall than VENSES++, its performance is still relatively low, e.g., the accuracy for the DNI v. INI classification is 55%. The authors argue that data sparseness is the biggest problem.

Also very closely related is Gerber and Chai (2010), which presents a study of implicit arguments for a group of frequent nominal predicates. Gerber and Chai (2010) model the task as a classical supervised task and implement a number of syntactic, semantic, and discourse features such as the sentence distance between an NI and its potential filler, their mutual information, and the discourse relation holding between the spans containing the target predicate and the potential filler.

While both Gerber and Chai (2010) and the SemEval-10 Task-10 deal with finding fillers for uninstantiated arguments, there are important differences between the two data sets, which make the results not directly comparable. Gerber and Chai’s corpus consists of newswire texts (Wall Street Journal), which is annotated with NomBank/PropBank roles. The data cover 10 nominal predicates from the commerce domain, with—on average—120 annotated instances per predicate. The Task-10 corpus consists of narrative texts annotated under the FrameNet paradigm. Crucially, this corpus provides annotations for running texts not for individual occurrences of selected target predicates. It thus treats many different general-language predicates of all parts of speech. While the overall size of the corpus in terms of sentences is comparable to Gerber and Chai’s corpus, the SemEval corpus contains many more target predicates and fewer instances for each.<sup>3</sup> These properties make it much harder to obtain good results on the SemEval corpus, which is supported by the fact that the NI resolution results obtained by the Task-10 participants are significantly below those

<sup>3</sup>E.g., Ruppenhofer et al. (2010) report that there are 1,703 frame instances covering 425 distinct frame types, which gives an average of 3.8 instances per frame.

reported by Gerber and Chai (2010).

While the SemEval-10 Task-10 is harder than the problem tackled by Gerber and Chai (2010), we also believe it is more realistic. Given the complexity of annotating semantic argument structures in general and null instantiations in particular, it seems infeasible to annotate large amounts of text with the required information. Hence, automated systems will always have to make do with scarce resources. We investigate different strategies of incorporating linguistic background knowledge to overcome this data sparseness problem, e.g., by explicitly modeling the DNI v. INI distinction, which is ignored by Gerber and Chai (2010). We also think that the task is best modeled as a semi-supervised task which combines the training data with FrameNet data not annotated for NIs.

Another line of research that is related to the goals of our effort is the work on zero pronoun resolution in pro-drop languages such as Japanese or Spanish. Iida et al. (2007) discuss the relevance of the semantic role labeling and zero-anaphora resolution tasks to each other and study how methods used in one task can help in the other. Still, their work is different from our task in two respects. First, it has a different coverage. Of the kinds of omissions that we consider to be null instantiations, Iida et al. (2007) target only the subset of constructionally licensed omissions. In addition, they seem to treat cases of co-instantiation or argument sharing—for instance subjects shared across conjoined VPs—as involving argument omission, which is not how similar cases would be treated in our FrameNet-style annotations. Second, in their system implementation Iida et al. (2007) use only syntactic patterns but no semantic information about the semantic class ( $\approx$  frame) of the predicate missing an argument or about the interrelations between the predicate missing an argument and the predicate(s) where coreferent mentions of the missing argument appear. Palomar et al. (2001) similarly use syntactic rather than semantic information in their work on Spanish, which only allows constructionally licensed subject omissions.

## 4 Data

In our experiments we used the corpus distributed for the SemEval-10 Task-10 on “Linking Events and Their Participants in Discourse” (Ruppenhofer et al., 2010). The data set consists of two

texts from Arthur Conan Doyle, “The Adventure of Wisteria Lodge”(1908) and “The Hound of the Baskervilles” (1901/02). From the first text, the second part entitled “The Tiger of San Pedro” (henceforth “Tiger”) was annotated and served as training data in the task; from the second text (henceforth “Hound”) chapters 13 and 14 were annotated and served as test data. The annotation consists of frame-semantic argument structure, coreference chains, and information about null instantiation, i.e., the NI type (DNI vs. INI) and the filler, if available in the text. Table 1 provides basic statistics about the data set.

In a qualitative analysis, we also considered a randomly chosen subset of 50 frame instances from the training data with at least one uninstantiated FE-set (see Section 6).

## 5 Modeling

We approach our three sub-decisions separately. The first sub-task, determining which, if any, frame elements are missing relies on information from the FrameNet release. Of particular importance is information about the three types of relationships between the core Frame elements: Core-Set, Excludes, and Requires. Given that we start with gold standard annotation of the overtly instantiated elements, we reason about the FE relations in the frame at issue to determine which FEs are to be considered as missing. For instance, consider the instance of the *Similarity* frame evoked by *different* in (10).

- (10) Falkner can be related to the “New South” literature but [his approach]<sub>Ent1</sub> was **different**<sub>Sim</sub>.

As discussed in Section 2, there are two FE-relation instances defined for the *Similarity* frame: a *Requires* relation between ENTITY\_1 and ENTITY\_2 and an *Excludes* relation between ENTITIES and ENTITY\_1 and ENTITY\_2. Given that ENTITY\_1 is instantiated, we conclude due to the *Excludes* relation that ENTITIES does not have to be treated as NI; given the *Requires* relation, we conclude that ENTITY\_2 does.

Our second sub-decision is to decide whether a frame element that we have found to be null-instantiated has an anaphoric (DNI) or an existential (INI) interpretation. Our approach for making this decision is the following. First, we check whether the omission we are looking at is licensed by a specific grammatical construction

which specifies the interpretation type of the argument it suppresses. For instance, we would treat the missing by-phrase agent of a passive as omitted with existential interpretation. Besides passive, we only consider imperatives at this point, although there are additional but less frequently occurring valence-suppressing constructions.

In our specific case of (10), there is no relevant construction that we can blame the omission on and we thus consider the omission to be lexically licensed. Since that is so, we next look at the FrameNet annotations for the specific frame evoking element. Either we only look at the annotations of the particular lexical unit that occurs in our text, or we consider statistics aggregated across all lexical units in a frame. In either case, for the frame element under consideration we choose that type of interpretation type that is more common in the annotated data. For *different* we find that uninstantiated cases of ENTITY\_2 are always labeled DNI and so in processing (10) we would choose DNI as well. Heuristics are needed when there either are no relevant annotations or when the frequencies of DNI and INI are tied.<sup>4</sup> The simplest heuristic is to simply choose one interpretation type as a default, which is what we do.

The final decision we have to make concerns uninstantiated FEs for which we have settled on the anaphoric interpretation type. For these, we have to locate, if possible, a coreferring antecedent mention. Any coreferring mention will do since we evaluate against coreference chains.<sup>5</sup> In theory, we could use customized strategies for antecedent finding depending, for instance, on whether the null instantiation is licensed by a construction or by a lexical item, or depending on the identity of the null-instantiated frame element. However, at the moment we treat the problem of antecedent finding in the same way for all null-instantiated frame elements.

One approach we pursue for identifying a suitable mention/chain relies on the semantic types that FrameNet specifies for frame elements. Specifically, we look up in FrameNet the semantic type(s) of the FE that is unexpressed. With that in-

<sup>4</sup>One might additionally choose to employ heuristics when the number of annotated instances is very small, or when the frequencies of DNI and INI are very close, though not tied. We have not used such heuristics here.

<sup>5</sup>Note that we have chains of length 1, since we for instance need to be able to reify whole sentences as referents that can be the antecedents for unexpressed MESSAGE, CONTENT or similar FEs of predicates such as *know* or *confess*.

data set	sentences	tokens	frame instances	frame types	overt frame elements	DNIs (resolved)	INIs
train	438	7,941	1,370	317	2,526	303 (245)	277
test	525	9,131	1,703	452	3,141	349 (259)	361

Table 1: Statistics for the SemEval-10 Task-10 corpus

formation in hand, we consider all the coreference chains that are active in some window of context, where being active means that one of the member mentions of the chain occurs in one of the context sentences. We try to find chains that share at least one semantic type with the FE in question. This is possible because for each chain, we have percolated the semantic types associated with any of their member mentions to the chain.<sup>6</sup> If multiple chains remain that are compatible with the FE in question, we select between them by some criterion. In particular, we prefer to link the FE to that chain that has the mention closest to the FE in question in terms of intervening leaf nodes.<sup>7</sup> If we find no chain at all within the window that has semantic types compatible with our FE, we guess that the FE has no antecedent.<sup>8</sup> Note also that in our current set-up we have defined the semantic type match to be a strict one. For instance, if our FE has the semantic type *Entity* and an active chain is of the type *Sentient*, we will not get a match even though the type *Sentient* is a descendant of *Entity* in the hierarchy in which semantic types are arranged.

## 6 Experiments

To gain a better understanding of our results for the full NI resolution task, we performed a qualitative analysis on a subset of 50 frames from the training set, in which one or more Frame elements were uninstantiated. We focus here on the first two sub-decisions that have to be made in the automatic analysis of null instantiations: which specific FEs should be treated as null-instantiated and

<sup>6</sup>In the official FrameNet database, not every frame element is assigned a semantic type. We modified our copy of FrameNet so that every FE does have a semantic type by simply looking up in WordNet the path from the name of a frame element to the synsets that FrameNet uses to define semantic types.

<sup>7</sup>Other criteria are easily conceivable. We might, for instance, use a tree-based distance measure, or link the FE to the chain that has the most mentions within the window of context.

<sup>8</sup>Alternatively, we could have widened the window of context in the hope of hitting upon a suitable chain.

which interpretation type the relevant FEs have.

The distribution of frames in this set was as follows: 33 frames occurring only once, 4 instances of *Arriving*, 3 instances of *Self-motion* and 2 of *Departing*. In 3 of the 6 instances of *Calendric\_unit* and in all 3 instances of *Self-motion*, our NI analysis system made errors. These are two challenging frames to handle which happen to be frequent in our data.

We also see that in our data, we have many nouns as frame evoking elements (FEEs). 28 of 50 FEEs are nouns, 15 verbs, and 7 adjectives. This distribution also contributes to an overall lower performance of our system because the error rate is highest for nouns, middling for adjectives, and lowest for verbs.<sup>9</sup> In our first system setting, where we use frame-level NI statistics and where we use INI as the default interpretation type when FrameNet either has no relevant data or shows equal probability for DNI and INI, the error rate on nouns is 53.6%, on adjectives 28.6%, and on verbs 13.3%.

In the first setting with INI as default, the system made no error on 31 of the 50 frames (62%). The 50 frame instances analyzed contain 62 FE-Sets that are not instantiated. (Recall that a single predicate may omit more than one argument at the same time.) Of these 62 sets, 38 are classified correctly as INI or DNI (61.3%) and the remaining 24 incorrectly. The predominant error type is the system positing INI where the gold value is DNI (16 of 24). The remaining errors are the other way around.

Given that for our data set, the baseline of guessing the DNI majority class is 52.2%, our system configuration has noticeably better precision at 62%. Importantly, we also have 100% recall for uninstantiated FE-sets unlike the systems in the SemEval task.

In our second NI analysis setting, we again use

<sup>9</sup>The same differences among the parts-of-speech can also be seen, for instance, in the performance on labeling of explicit FEs where the treatment of verbal predicators is more successful.

	Accuracy
Maj. Baseline	52.2%
PerFrame	61.3%
PerLU	66.0%

Table 2: Distinguishing DNIs and INIs

INI as the default value but we use lexical unit-specific NI-statistics rather than aggregate statistics over all lexical units in the frame. Doing so improves the result a bit: we classify 41 of 62 FE-sets (66%) correctly, for a 4.7% improvement over the previous setting. Table 2 provides a summary of the results.

Finally, we look at the sources of error for our first setting. As noted above, there were 19 frame instances where at least one FE-set was classified incorrectly. The main reasons for these errors were:

- With 5 frame instances, the error results because the aggregate frame-level statistics are distorted. This is due to two reasons: there are few annotated instances, or a “deviant” lexical unit is overrepresented.
- In another 5 frame instances, the use of INI as a default is inappropriate. These are cases where either no lexical unit in the frame is annotated at all, or where the frame was created and annotated before the practice of annotating missing arguments was adopted.
- In 4 frame instances, a misclassification occurs because the instance of the frame in our test data occurs in a special linguistic context that overrides the majority interpretation type that can be observed in the FrameNet data. For instance, the context in our data may be generic, while the majority of cases in FrameNet annotations are episodic.
- 4 frame instances belong to the linguistically difficult frames where the gold standard analysis itself may not be fully worked out. A good example of this is Calendric unit.

While our manually inspected data set is small, it seems we must conclude from this qualitative analysis that even a reasonable, linguistically motivated use of the available FrameNet data won’t yield the correct result for NI-classification in all

cases. One difficulty arises from FrameNet’s annotation practice, which does not select instances randomly. Hence, the statistics about indefinite v. definite interpretations for a given FE that can be gleaned from FrameNet are not necessarily accurate. At this point, we do not know the exact number of frames where, for instance, a skew in the annotated LUs or the annotated instances of a particular LU would lead to incorrect classifications. But even if FrameNet had annotated a large number of randomly chosen instances for all LUs, our current system would not achieve perfect performance because it lacks a way of detecting constructions and contexts (such as generic or habitual sentences) that can override the majority interpretation type. Complementing our system with an additional analysis step which attempts to identify different event types thus seems beneficial. The work by Reiter and Frank (2010) and Mathew and Katz (2009) on generic NPs and sentences could be a starting point.

Since there are only very few resolved NIs in the 50 frame data set we used to evaluate the first two sub-tasks, we evaluated the NI resolution task (i.e., the third sub-task) on the whole SemEval-2010 Task-10 test set. We employed the best performing SemEval system, SEMAFOR (Chen et al., 2010), as a baseline. Even though our NI resolution strategy is still fairly basic, taking only the semantic type of potential fillers into account, our system reduces the resolution errors for the complete pipeline by 14% compared to SEMAFOR. This may be due to the fact that our DNI v. INI classification is better. As the DNI v. INI distinction was not evaluated for the shared task, we cannot directly compare our results on this sub-task against SEMAFOR. However, Chen et al. (2010) provide a confusion matrix for argument classification (Table 3 in their paper), which suggests that only 3% of DNIs are correctly identified. The majority of unidentified DNIs are misclassified as INIs (52%).

SEMAFOR is, however, a bit better at identifying the correct boundaries for correctly found antecedents (100% NI linking overlap v. 89% for our system). The reason for this may be that we consider more varied antecedents. In particular, we also consider full sentence antecedents. Example (11) illustrates the problem of identifying the correct boundaries for full sentence antecedents. The gold annotation identifies both (a)

and (b) as the antecedent of the CONTENT FE of the `Experiencer_focus` frame evoked by *pleasure* in (c), while our system resolved the NI only to (b).

- (11) a. "I must congratulate you, Inspector, on handling so distinctive and instructive a case.  
b. Your powers, if I may say so without offence, seem superior to your opportunities."  
c. Inspector Baynes's small eyes twinkled with **pleasure**<sub>Exp.foc</sub>.

## 7 Conclusion

We presented a novel approach to recognizing and resolving null instantiations. We split the task in three sub-tasks: identification of NIs, distinguishing definite and indefinite NIs, and resolving NIs to a suitable referent in the text. We paid particular attention to the first two sub-tasks. The first task was addressed by making use of background knowledge about interdependencies between frame elements. For the second task, we employed a hybrid system which combined rules for identifying syntactic constructions with statistics about DNI v. INI distributions for different lexical units or frames. For the resolution task we made use of FrameNet's semantic type information for frame elements which we enriched with semantic information from WordNet.

We showed that our system has a noticeably better performance on the whole pipeline than the best system participating in the SemEval-10 NI resolution task. This is probably due to the fact that we employ a more sophisticated system for identifying DNIs.

However, an error analysis revealed that there are also areas where our system could be improved. Obtaining reliable statistics for lexically licensed NIs from FrameNet proves difficult because FrameNet data were not randomly selected. It may be possible to overcome this shortcoming by trying to glean information about NIs from unannotated data, e.g., by using semantic similarity to cluster syntactic arguments. A preprocessing component which identifies different event types (generics, habituals etc.) might also help to identify DNIs in a more reliable fashion. Furthermore, our strategy for finding antecedents is still fairly basic. Adding additional features, e.g., along the lines of Gerber and Chai (2010) will probably lead to better performance.

## Acknowledgments

This research has been funded by the German Research

Foundation DFG (MMCI Cluster of Excellence and grant PI 154/9-3).

## References

- D. Chen, N. Schneider, D. Das, N. A. Smith. 2010. SEMAFOR: Frame Argument Resolution with Log-Linear Models. In *Proc. of SemEval-2010*, 264–267.
- D. Das, N. Schneider, D. Chen, N. A. Smith. 2010. Probabilistic Frame-semantic Parsing. In *Proc. of NAACL-HLT-10*, 948–956.
- R. Delmonte. 2008. *Computational Linguistic Text Processing: Lexicon, Grammar, Parsing and Anaphora Resolution*. Nova Science, New York.
- C. Fillmore. 1986. Pragmatically Controlled Zero Anaphora. In *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, 95–107.
- M. Gerber, J. Y. Chai. 2010. Beyond NomBank: A Study of Implicit Arguments for Nominal Predicates. In *Proc. of ACL-2010*, 1583–1592.
- R. Iida, K. Inui, Y. Matsumoto. 2007. Zero-anaphora Resolution by Learning Rich Syntactic Pattern Features. *ACM Transactions on Asian Language Information Processing (TALIP)*, 6:1:1–1:22.
- H. Liu, P. Singh. 2004. ConceptNet: A Practical Commonsense Reasoning Toolkit. *BT Technology Journal*, 22(4):211–226.
- T. Mathew, G. Katz. 2009. Supervised categorization of habitual and episodic sentences. In *Sixth Midwest Computational Linguistics Colloquium*, Bloomington, Indiana.
- M. Palomar, L. Moreno, J. Peral, R. Muñoz, A. Ferrández, P. Martínez-Barco, M. Saiz-Noeda. 2001. An Algorithm for Anaphora Resolution in Spanish Texts. *Computational Linguistics*, 27:545–567.
- N. Reiter, A. Frank. 2010. Identifying Generic Noun Phrases. In *Proc. of ACL-10*, 40–49.
- J. Ruppenhofer, M. Ellsworth, M. R. L. Petruck, C. R. Johnson, J. Scheffczyk. 2006. FrameNet II: Extended Theory and Practice. available at [http://framenet.icsi.berkeley.edu/index.php?option=com\\_wrapper&Itemid=126](http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126).
- J. Ruppenhofer, C. Sporleder, R. Morante, C. Baker, M. Palmer. 2010. SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. In *Proc. of SemEval-2010*, 45–50.
- S. Tonelli, R. Delmonte. 2010. VENSES++: Adapting a Deep Semantic Processing System to the Identification of Null Instantiations. In *Proc. of SemEval-2010*, 296–299.