Combining word sense and usage for modeling frame semantics

Diego De Cao†, Danilo Croce†, Marco Pennacchiotti‡, Roberto Basili†

(†) DISP, University of Rome Tor Vergata, Roma (Italy), {decao,croce,basili}@info.uniroma2.it

(‡) Computational Linguistics Saarland University Saarbrücken, Germany pennacchiotti@coli-uni sb.de

Abstract

Models of lexical semantics are core paradigms in most NLP applications, such as dialogue, information extraction and document understanding. Unfortunately, the coverage of currently available resources (e.g. FrameNet) is still unsatisfactory. This paper presents a largely applicable approach for extending frame semantic resources, combining word sense information derived from WordNet and corpus-based distributional information. We report a large scale evaluation over the English FrameNet, and results on extending FrameNet to the Italian language, as the basis of the development of a full FrameNet for Italian.

1 Introduction and Related Work

Models of lexical meaning are explicit or implicit basic components of any text processing system devoted to information extraction, question answering or dialogue. Several paradigms proposed for a variety of notions, such as word sense (Miller et al., 1990) or frame semantics (Baker et al., 1998), have given rise to large scale resources, respectively WordNet and FrameNet. Recent studies (e.g. (Shen and Lapata, 2007)) show that the use of FrameNet can potentially improve the performance of Question Answering systems. Yet, (Shen and Lapata, 2007) also point out that the low coverage of the current version of FrameNet significantly limits the expected boost in performance. Other studies have shown similar evidences for Recognizing Textual Entailment (RTE) (Clark et al., 2007; Burchardt et al., 2008): most examples of the RTE challenges corpora can be solved at the predicate-argument structure level, but FrameNet coverage is still a major problem.

Approaches to (semi-)automatically acquire frame information are then today a priority to solve these problems. Despite this, not many efforts have been paid so far in this direction. (Burchardt et al., 2005) presented Detour, a system for predicting frame assignment of potential lexical units not covered by FrameNet, by using the paradigmatic information enclosed in WordNet. Although the authors do not fully solve the problem related to the fuzzy relationships between senses and frames, they propose an empirical association measure for ranking frame candidates according to sense information as stored in WordNet. To our knowledge, this is the only work trying to bridge frame membership to referential properties of lexical senses. (Pitel, 2006) presents a preliminary study on the applicability of semantic spaces and space geometrical transformations (namely, Latent Semantic Analysis) to expand FrameNet, but the investigation is too limited in scope to draw relevant conclusions. Finally, (Padó et al., 2008) propose a method to automatically label unknown semantic roles of event nominalizations in FrameNet, but their method needs a large amount of annotated verbal data.

Another important limitation of FrameNet is the limited support to multilinguality, which is becoming a critical issue in real NLP applications. In recent years, some efforts have focused on the manual adaptation of the English FrameNet to other languages (e.g., German (Burchardt et al., 2006) and Spanish (Subirats and Petruck, 2003)). Unlike PropBank, FrameNet is in fact suitable to cross-lingual induction, as frames are mostly defined at the conceptual level, thus allowing cross-lingual interpretation. Yet, all these efforts consist in manually defining frame linguistic knowledge (e.g. lexical units) in the specific language, and in annotating a large corpus, thus requiring a large human effort. While attempts to automate the annotation process are quite promising (Pado and Lapata, 2007), they require the availability of a parallel corpus, and leave open the issue of inducing the resource as a whole in a new language.

In this work, we investigate novel methods for automatically expanding the English FrameNet, and sup-
porting the creation of new ones in other languages (namely Italian), thus tackling the abovementioned problems of coverage and multilinguality. The proposed methods are inspired by the basic hypothesis that FrameNet can be automatically modeled by a fruitful interaction between advanced distributional techniques, and paradigmatic properties derived from WordNet.

In particular, in this paper we focus on the application of such methods to study the semantics of the core elements of FrameNet, i.e. the lexical units (hereafter LUs). Lexical units are predicates (nouns, verbs, adjectives, etc.) that linguistically express the situation described by a frame. Lexical units of the same frame share semantic arguments. For example the frame KILLING has lexical units such as: assassin, blood-bath, fatal, massacre, kill, suicide. These predicates share semantic arguments such as KILLER, INSTRUMENT and VICTIM. Our goal is to combine corpus distributional evidence with WordNet information to supply three tasks: the induction of new LUs not already in FrameNet (unknown LUs); the reduction of LUs polysemy by mapping them to WordNet synsets; the translation of English LUs into Italian.

The paper is organized as follows. In Section 2 we describe our FrameNet paradigmatic and distributional model, and we discuss how these two models can be combined in a single framework. In Section 3 we analyze the applicability of these models to the three proposed experimental tasks, and discuss the results. Finally, in Section 4 we draw final conclusions and outline future work.

2 Paradigmatic and distributional models of frame semantics

In this section we describe our paradigmatic, distributional and combined models for representing FrameNet. The general goal of each of these three methods is to offer a computational model of FrameNet. In such a model, frames and LUs should have a specific computational representation (e.g. vectors), and allow the computation of similarity either among different LUs or between a frame and a LU. Such model thus offers explicit means to use FrameNet in a NLP task or to expand FrameNet, e.g. by assigning unknown LUs to its most similar frame, or by mapping a LU to its proper WordNet synset(s). A key notion for these tasks is the definition of a principled and reliable semantic similarity measure sim to be applied to frames and LUs.

2.1 Paradigmatic model

The basic intuition behind our paradigmatic model is that knowledge about predicates of a frame, through a (possibly limited) set of LUs, allows to detect the set of the suitable WordNet senses able to evoke the same frame. These senses are topologically related to (one or more) sub-hierarchies capturing the lexical semantics implicit in the frame. We propose a weakly-supervised approach to discover these structures. The main idea is that frames correspond to specific sub-graphs of the WordNet hyponymy hierarchy, so that these latter can be used to predict frames valid for other LUs, not yet coded in FrameNet. Figure 1 reports the WordNet sub-hierarchy covering the frame PEOPLE_BY_AGE: here, the frame’s nominal LUs {adult, adolescent, baby, boy, infant, kid, geezer, teenager, youngster, youth} are all represented with the senses correctly referring to the frame. The correct senses (e.g. sense 1 of youth out of its 6 potential senses) are selected as they share most specific generalizations with the other LUs. This graph can be intended as an “explanation” of the lexical semantic properties characterizing the frame: future predictions about new LUs can be done on the basis of the graph as a paradigmatic model for PEOPLE_BY_AGE. We call such a graph the WordNet model of the frame. As WordNet organizes nouns, verbs and other parts-of-speech in different hierarchies, three independent WordNet models (one for each part-of-speech) are created for each frame.

Formally, given the set \( F \) of the LUs of a frame, a WordNet model is built around the subset \( S_F \) of WordNet synsets able to generalize the largest number of words in \( F \). The WordNet model \( WN_F(\Gamma, W) \)

of a frame \( F \), is a graph

\[
WN_F(\Gamma, W) = < W, S_F, L_F, h, sim_{WN}, m > \tag{1}
\]

In the following, we will use the same notation for a frame \( F \) and for the set of its known lexical units, as in our approach we use LU membership as a basic definition of a frame.
where: $W \subseteq F$ are the subset of all LUs in $F$ having same part-of-speech $\Gamma \in \{\text{verb}, \text{noun}, \text{adjective}\}$, $S_F$ is the subset of synsets in WN needed to generalize words $w \in W$; $L_F \subseteq S_F$ are the lexical senses of $w \in W$ subsumed by $S_F$; $h \subseteq S_F \times S_F$ is the projection of the hyponymy relation of WN in $S_F$; $m \subseteq W \times 2^{S_F}$ is the lexical relation between words $w \in W$ and synsets in $L_F$; $\sim_{WN} : S_F \rightarrow \mathbb{R}$ is a weighting function that expresses the relevance of each sense $\sigma \in S_F$ for the frame, as it is represented by its words in $F$.

The model exemplified in Figure 1 is $\text{WN}_{\text{People,by,Age}}(\text{noun}, \{\text{adult, ..., youth}\})$, where $L_F = \{\text{adult#1, adolescent#1, baby#1, boy#1, boy#2, boy#3, ..., youth#1}\}$ and the set $S_F$ corresponds to the sub-hierarchies dominated by the synsets #6026, #9622621, #9285271 and #9015843.

The overall goal of computing the WordNet model is to determine the similarity function $\sim_{WN} : S_F \rightarrow \mathbb{R}$, expressing the relevance of a synset $\sigma \in S_F$ as a good representative of a frame $F$. This is what is hereafter referred to as the paradigmatic similarity model between word senses and frames.

### 2.1.1 Paradigmatic Similarity measures

Given the WordNet hierarchy separation on part-of-speeches, the similarity function $\sim_{WN}$ is independently defined for verbs, nouns and adjectives.

For nouns we adopt conceptual density (cd) (Agirre and Rigau, 1996; Basili et al., 2004), a semantic similarity measure defined for word sense disambiguation tasks. The cd score for a sense $\sigma \in S_F$ is the density of the WordNet sub-hierarchy rooted at $\sigma$ in representing the set of nouns in $F$. The intuition behind this model is that the larger the number of all and only LUs in $F$ that are generalized by $\sigma$, the better it captures the lexical semantics intended by the frame. Coarse generalizations (i.e. synsets higher in the hierarchy) are penalized, as they give rise to bushy hierarchies, covering too many words not in the target $F$. The greedy algorithm proposed in (Basili et al., 2004) selects the subset of synsets able to “cover” (i.e. generalize) all the input words and characterized by the highest cd values. The set of such synsets and their corresponding sub-hierarchies forms a graph derived from a set of LUs $F$. The result is the WordNet model $\text{WN}_F$ for $F$, i.e. the minimal subset of WordNet that explains all (the possible) LUs in $F$ with the maximally similar senses.

Figure 1 shows that correct senses (e.g. the sense 1 of youth out of the 6 potential senses) are generally detected and preserved in the model. Irrelevant senses that do not share any common hypernym with other words in $F$ are neglected. Conceptual density scores can be used to rank individual senses as in the case of boy.

Given a frame $F$, the above model can be naturally used to compute the similarity between a noun $n \notin F$ and $F$. This is particularly useful in LU induction task, as described in Section 3.1. To do so, the similarity $\sim_{WN}(F, n)$ between $n$ and $F$ is derived by computing the cd scores over the set $F \cup \{n\}$. The $\sim_{WN}(F, n)$ is the maximal cd of any synset $\sigma_n \in S_F$ that is also hypernym of a lexical sense of $n$. In the example, the noun boy would receive a score of 0.117 through the hypernym \{child, kid\}, according to its third sense in WordNet 2.0 (“\{son, boy\}”).

As conceptual density can be only applied to nouns, when verbs $v$ are considered, we exploit the synonymy and co-hyponymy relations. The following similarity $\sim_{WN}(F, v)$ is computed:

$$\sim_{WN}(F, v) = \begin{cases} 1 \text{ if } \exists K \subset F \text{ such that } |K| > \tau \text{ AND } \\ \forall w \in K \text{ w is a co-hyponym of } v \\ \epsilon \text{ otherwise} \end{cases}$$

(2)

For adjectives, the similarity $\sim_{WN}(F, a)$, is computed on the basis of the synonymy relation, as follows:

$$\sim_{WN}(F, a) = \begin{cases} 1 \text{ if } \exists w \in F \text{ such that } \\ w \text{ is a synonym of } tw \\ \epsilon \text{ otherwise} \end{cases}$$

(3)

The overall model $\text{WN}_F$ is used to predict if a frame $F$ is a correct situation for a given unknown LU $ul \notin F$ (a noun, a verb or an adjective), whenever $\sim_{WN}(F, ul) > \epsilon$. This can be used as a frame predictor for a $ul$ currently not foreseen in the Berkley database but possibly very frequent in a specific corpus, as described in Section 3.1.

### 2.2 Distributional model

The distributional model is based on the intuition that FrameNet frames and LUs can be modelled in a semantic space, where they are represented as distributional co-occurrence vectors computed over a corpus. Such framework, it is possible to compute the similarity between a LU and a frame, by evaluating the distance of their vectors in the space.

Semantic spaces have been widely applied in several NLP tasks, ranging from information retrieval to paraphrase rules extraction (Lin and Pantel, 2001). The intuition is that the meaning of a word can be described by the set of textual contexts in which it appears (Distributional Hypothesis (Harris, 1964)), and that words with similar vectors are semantically
related. This distributional approach has been often claimed to support the language in use view on meaning. Word space models (Schütze, 1993) have been shown to emphasize different aspects of lexical semantics: associative (i.e. topical) information between words, as well as paradigmatic information (i.e. in absentia) or syntagmatic information (i.e. in presentia).

In our setting, the goal is to leverage semantic spaces to capture the notion of frame – i.e. the property of “being characteristic of a frame”. To do so, we model a lexical unit \( l \) as a vector \( \vec{l} \), whose dimensions represent the set of contexts of the semantic space. In our space, contexts are words appearing in a \( n \)-window of the lexical unit: such a space models a generic notion of semantic relatedness – i.e. two LUs close in the space are likely to be either in paradigmatic or syntagmatic relation (Pado, 2007; Sahlgren, 2006). The overall semantic space is then represented by a matrix \( M \), whose rows describe LUs and whose columns describe contexts.

We reduce in dimensionality the matrix \( M \) by applying Singular Value Decomposition (SVD) (Landauer and Dumais, 1997), a decomposition process that creates an approximation of the original matrix, aiming to capture semantic dependencies between source vectors, i.e. contexts. The original space is replaced by a lower dimensional space \( M_k \), called \( k \)-space in which each dimension is a derived concept. The matrix \( M \) is transformed in the product of three new matrices: \( U \), \( S \), and \( V \) such that \( M = USV^T \). Truncating \( M \) to its first \( k \) dimensions means neglecting the least meaningful dimensions according to the original distribution. \( M_k \) captures the same statistical information in a new \( k \)-dimensional space, where each dimension is a linear combination of some original features. These newly derived features may be thought of as artificial concepts, each one representing an emerging meaning component as a linear combination of many different words (or contexts).

The SVD reduction has two main advantages. First, the overall computational cost of the model is reduced, as similarities are computed on a space with much fewer dimensions. Secondly, it allows to capture second-order relations among LUs, thus improving the quality of the similarity measure.

Once the vectors \( \vec{l} \) for all FrameNet LUs are available in the reduced space, it is also possible to derive a vectorial representation \( \vec{F} \) of a whole frame \( F \). Intuitively, \( \vec{F} \) should be computed as the geometric centroid of the vectors of its lexical units. Unfortunately, such a simple approach is prone to errors due to the semantic nature of frames. Indeed, even if the LUs of a given frame describe the same particular situation, they can typically do that in different type of contexts. For example, the LUs assassinate and holocaust evoke the KILLING frame, but are likely to appear in very different linguistic contexts. Then, the vectors of the two words are likely to be distant in the space. Consequently, different regions of the semantic space may act as good representations for the same frame: these regions corresponds to clusters of LUs which appear in similar contexts (e.g. \{holocaust, extermination, genocide\} and \{suicide, euthanasia\}).

We then adopt a clustering approach to model frames: each frame is represented by the set of clusters \( C_F \) of its lexical units. Clusters \( C_F \) are composed by lexical units close in the space and can have different size. They are computed by using an adaptive (unsupervised) algorithm, based on \( k \)-means (Heyer et al., 1999; Basili et al., 2007), applied to all known LUs of \( F \). Each cluster \( c \in C_F \) is represented in the space by a vector \( \vec{c} \), computed as the geometric centroid of all its lexical units.

In this framework, it is then possible to compute the similarity between an unknown LU \( ul \) and a frame \( F \), as the the cosine distance between the vector \( \vec{ul} \) and the closest centroid \( c \in C_F \):

\[
sim_{LSF}(F, ul) = \text{argmax}_{c \in C_F} \text{sim}_{cos}(\vec{ul}, \vec{c})
\]

(4)

Given this measure, it is finally possible to assign an unknown \( ul \) to one or more of the most similar frames.

2.3 Combining paradigmatic and distributional models

In order to be effective in a NLP task, a model of lexical meaning should typically account for both the paradigmatic and distributional similarity. The following definition thus hold:

\[
\mu(F, w) = \Psi(D(w, F), P(w, F))
\]

(5)

where \( \mu(F, w) \) is the association between a word \( w \) and a frame \( F \), \( \Psi \) is a composition operator applied to the corpus-driven distributional measure \( D(w, F) \) and to the paradigmatic similarity \( P(F, w) \). Notice that in this work \( \text{sim}_{LSF}(F, w) \) and \( \text{sim}_{MN}(F, w) \) are used as models of \( D(F, w) \) and \( P(F, W) \), respectively. Different combinations \( \Psi \) are here possible, from simple algebraic operations (e.g. linear combinations) to more complex algorithmics. We will explore this latter issue in section 3.1 where the evaluation of a combined model for LU induction is reported.

3 Experiments

In this section we experiment our proposed models on three different tasks: induction of new LUs (Sec-
tion 3.1), mapping LUs to WordNet synsets (Section 3.2), and automatic acquisition of LUs in Italian (Section 3.3). In all the experiments we use FrameNet 1.3, consisting of 795 frames and about 10,196 LUs (7,522 unique LUs), as source information and as a gold standard. As regards WordNet, we adopt version 2.0, with all mappings from 1.6 applied through the Italian component of MultiWordNet (Pianta et al., 2008).

For computing vectors in the distributional model, we use the TREC-2002 Vol.2 corpus, consisting of about 110 million words for English. The contexts for the description of LUs are obtained as ±5 windows around each individual LU occurrence: each word occurring in this windows is retained as a potential context. A resulting set of about 30,000 contexts (i.e. individual words) has been obtained. The vector representing an individual LU is derived by computing pointwise mutual information between the LU and each context. The SVD reduction has been run over the resulting $7,522 \times 30,000$ matrix, with a dimension cut of $k = 50$, other values resulting in non-statistically different outcomes.

Experiments for Italian are run against the Italian component of the Europarl warrant corpus (Koehn, 2002), made of about 1 million sentences for about 36 millions tokens, for which about 87,000 contexts are used for the targeted LUs. Also for Italian a dimension cut of $k = 50$ has been applied.

3.1 Lexical Unit induction

The goal of this experiment is to tackle the FrameNet low coverage problem, by checking if our models are good in expanding FrameNet with new LUs. Formally, we define LU induction as the task of assigning a generic unknown lexical unit $ul$ not yet present in the FrameNet database to the correct frame(s). As the number of frames is very large, the task is intuitively hard to solve. A further complexity regards multiple assignments. Lexical units are sometimes ambiguous and can then be mapped to more than one frame (for example the word tea could map both to FOOD and SOCIAL_EVENT). Also, even unambiguous words can be assigned to more than one frame – e.g. child maps to both KINSHIP and PEOPLE_BY_AGE.

In the experiment, we simulate the induction task by executing a leave-one-out procedure over the set of existing FrameNet LUs, as follows. First, we remove a LU from all its original frames. Then, we ask our model to reassign it to the most similar frame(s), according to its similarity measure. We repeat this procedure for all lexical units and compute the accuracy in the assignment.

We experiment all three models: distributional, paradigmatic and the combined one. In particular, the combined model is applied as follows. First, for each frame $F$ we create its cluster set $C_F$ in the LSA space. Then, at each iteration of the leave-one-out a different LU $ul$ is removed from FrameNet, and the following steps are performed:

- We recompute the clusters for all frames $F_{ul}$ which contain $ul$, by neglecting $ul$.
- We compute the similarity $sim_{LFS}(F, ul)$ between $ul$ and all frames. During the computation we empirically impose a threshold: if a cluster $c \in C$ has a similarity $cos(c, ul) < 0.1$ (i.e. poorly similar to $ul$), it is neglected. Finally, all suggested frames are ranked according to $sim_{LFS}(F, ul)$.
- For each frame $F$ we also compute the similarity $sim_{WN}(F, ul)$ according to the paradigmatic model, by neglecting $ul$ in the computation of the WordNet model of each frame.
- We combine the distributional and paradigmatic similarities following the general Equation 5, by applying the following specific equation:

$$sim(F, ul) = sim_{LFS}(F, ul) \cdot sim_{WN}(F, ul)$$

Note, that in practice $sim(F, ul)$ acts as a re-ranking function of the previously obtained clusters $C_F$.

- We execute LU induction, by mapping $ul$ to the most similar $k$ frames according to $sim(F, ul)$.

3.1.1 Evaluation

We evaluate the model by computing the accuracy over the FrameNet gold standard. Accuracy is defined as the fraction of LUs that are correctly re-assigned to the original frame during the leave-one-out. Accuracy is computed at different levels $k$: a LU is correctly assigned if its gold standard frame appears among the best-$k$ frames ranked by the model. We experimented both on English (using FrameNet version 1.3), and on Italian. Since an Italian FrameNet is not available, we manually created a gold standard of 11 frames. Overall statistics on the data are reported in Table 1: the number of frames and LUs analysed is slightly reduced wrt FrameNet as we ignored the predicate

---

3For all occurrences of feature words, the POS tag has been neglected in order to verify the applicability of the model even with a shallow preprocessing. Words occurring less than 10 times in all windows are neglected in our experiments.

4Note that all appearances of $ul$ in the database are neglected (irrespectively from its POS tag, e.g. march as a verb vs. march as a noun)
Table 1: The Gold Standard for the test over English and Italian

| Language | Number of frames: | Number of LUs: |
|----------|-------------------|----------------|
| English  | 220               | 5042           |
|          | Most likely frames: Self_Motion (p=0.015), Clothing (p=0.014) |
| Italian  | 10                | 112            |
|          | Frames: Buildings, Clothing, Killing, Kinship, Make_noise, Medical_conditions, Natural_Features, Possession, Self_Motion, Text |

words absent from the targeted corpus (e.g. *moo* in *MAKE_NOISE*) and multiwords expressions as it was not possible to locate them unambiguously in the corpus (e.g. *shell out* in *COMMERCE_PAY*). Also, in order to get reliable distributional statistics, we filter out LUs occurring less than 50 times in the corpus, and frames with less than 10 LUs.

For all the experiments, the parameter $\tau$ in the Equation 2 of the paradigmatic model has been set to 2. As a baseline, we adopt a model predicting as best-$k$ frames the most likely ones in FrameNet – i.e. those containing the highest number of LUs. Results for English are reported in Figure 2. As shown, all methods improve significantly the baseline whereas accuracy values naturally improve along increasing values for $k$. The performance of the paradigmatic model are significantly high even for very small $k$. The best model is given by the combination of distributional and paradigmatic similarity, producing significant improvements wrt the paradigmatic model alone.

Results for Italian are reported in Figure 3. The leave-one-out test has been applied as for English, but over a manually compiled set of 527 LUs for the 11 frames used as gold standard. These LUs have been obtained via direct translation of the English Framenet LUs. In order to evaluate LUs for which a consistent distributional model was available, only those occurring at least 50 times in the Europarlament corpus have been selected: this amounts to a total number of 112 Italian LUs. The paradigmatic model for the test has been obtained using as source the LUs in the English FrameNet. As the computation of the $\text{sim}_{WN}(F, w)$ depends only on the hyponymy hierarchy, for each Italian noun $n$ the conceptual density computation over the set $\{n\} \cup F$ is applied, where $F$ is given by the LUs in English. The interlingual index is here used to map every $n$ to its lexical senses (i.e. synsets) in the English WN. Then, the computation of the greedy algorithm is applied exactly as in the monolingual process. The same approach has been used for verbs (Equation 2) and adjectives (Equation 3).

Although the limited scale of the experiment (only 11 frames are targeted), the evidence are similar as for the test over English: the combined model is always superior to the individual ones. High levels of accuracy are achieved, although the “most likely frame” baseline is much higher than for the English test. Similar trends are also observed for the paradigmatic model, reaching a plateau for smaller values of $k$. Overall results indicate that reliable predictions can be obtained for unknown LUs also when a whole Italian FrameNet is not yet available. Our method can then be used to support lexicographers in the task of building a new FrameNet, in the specific stage of adding LUs to frames.

Results suggest that the WordNet models derived from the English LUs are valid predictors also for Italian words, as confirmed by the experiments in the next sections.

### 3.2 Assessing WordNet models of Frames

The goal of the experiment is to validate the notion of WordNet model of a frame as derived through the method discussed in Section 2.1. Formally, given the set of all possible WordNet senses $S_l$ of a given LU $l$, we aim at mapping each sense $s \in S_l$ to the correct frame $f \in F_l$, where $F_l$ is the set of frames in which $l$ appears. If a frame cannot be found for a given sense, the sense is simply neglected.

For example, the LU *burn* has 15 senses in WordNet and it belongs to 3 frames: *EMOTION_HEAT*, *EXPERIENCE_BODILY_HARM* and *PERCEPTION_BODY*. Figure 2 reports some of the possible correct mapping between senses and frames. Other senses, such as “destroy by fire” cannot be mapped to any existing frame.

By creating such an automatic mapping we achieve three goals. First, we disambiguate FrameNet lexical units. Second, we enrich WordNet synsets with new
Table 3: Statistics on nominal and verbal senses in the paradigmatic model of the English FrameNet

|                           | Nouns | Verbs |
|---------------------------|-------|-------|
| Targeted Frames           | 220   | 220   |
| Involved LUs              | 2,200 | 2,180 |
| Average LUs per frame     | 10.0  | 9.91  |
| LUs covered by WordNet    | 2,187 | 2,169 |
| Number of Evoked Senses   | 7,443 | 11,489|
| Average Polysemy          | 3.62  | 3.62  |
| Represented words (i.e. \(\sum_{F} W_{F}\)) | 2,145 | 1,270 |
| Average represented LUs   | 9.94  | 9.85  |
| Active Lexical Senses \((L_F)\) | 3,095 | 2,282 |
| Average Active Lexical Senses \((\|L_F\|/\|W_F\|)\) per word over frames | 1.27 | 1.79 |
| Active synsets \((S_F)\)   | 3,512 | 2,718 |
| Average Active synsets \((|S_F|/|W_F|)\) per word over frames | 1.51 | 2.19 |

In this experiment we focus on verbs and nouns, since they are core predicates expressing the targeted situation in sentences.
Each of these translated LU $l$ is a cross-lingual synonym of at least a sense in $S_F$ and is a candidate LUs for the frame in the new language, since it satisfies $\text{sim}_{WN}(F, l) > \epsilon$.

### 3.3.1 Evaluation

In the experiment we focus on Italian, for which a full FrameNet is not yet available, though a manual building process is currently underway (Tonelli and Pianta, 2008). As interlingual index we adopt the Italian component of MultiWordNet (Pianta et al., 2008). As shown in Table 4, the WordNet model allows to generate approximately 15,000 Italian LUs, partitioned in 6,600 nouns, 8,300 verbs and 130 adjectives.

| Type       | Number of LUs |
|------------|---------------|
| Total      | 15072         |
| Nouns      | 6611          |
| Verbs      | 8332          |
| Adjectives | 129           |

Table 4: Number of generated Lexical Units

To evaluate the quality of the translated LUs we performed two different tests. In the first test, we collected the 776 most frequent words in the Europarliament corpus, including many generic nouns and verbs, such as produrre (to produce/make), fare (to make/fabricate), avere (to have). Then we manually validated all the 1,500 system decisions regarding these words. A decision is accepted if the frame suggested for the word is correct for at least one of its senses. Accuracy is computed as the percentage of the correct system decisions over the number of validated cases. For some words no frame was predicted, as they were not in Wordnet or as no Wordnet model was able to correctly generalize them. The percentage of words receiving at least one correct prediction, i.e. assigned to at least one frame accepted by the annotators, is here called Coverage, and reported with the accuracy scores in the second line of Table 5. The above test was repeated also for more specific words, with a number of occurrences in the corpus ranging from 100 to 200. Results are reported in the third line.
Table 6: Excerpt of the Italian LUs in four different frames

| FRAME       | FRAME definition                                                                 | Italian LUs                                |
|-------------|----------------------------------------------------------------------------------|--------------------------------------------|
| BUILDINGS   | Words which name permanent fixed structures forming an enclosure and providing protection from the elements | autorimessa, cuccia, casolare, casotto, dependance, masseria, palazzina, ... |
| CLOTHING    | Anything that people conventionally wear                                          | cappotto, calzetta, camicia_da_notte, duepezzi, ... |
| SELF-MOTION | The Self mover, a living being, moves under its own power in a directed fashion.  | annaspare, arrancare, buttarsi, claudi-care, giro, ... |
| TEXT        | An entity that contains linguistic, symbolic information on a Topic, created by an Author at the Time of creation. | arringa, articolo_di_fondo, canzonetta, conto, poliza, vademecum, ... |

Table 5: Manual validation of the italian LUs generated through WordNet

| Frequency Range | Numb. of Test pairs | Numb. of Words | Acc. | Cov.  |
|-----------------|---------------------|----------------|------|-------|
| [722;55,000]    | 1,500               | 776            | 0.79 | 93%   |
| [100;200]       | 558                 | 357            | 0.87 | 94.3% |

of Table 5. These outcomes are surprisingly good especially considering that the computation of the individual $\text{sim}_{WN}(F, l)$ scores is fully automatic.

In a second test, the results generated by our method were compared against the words of the oracle manually developed for the experiment in Section 3.1. In this case, the predictions of the method about frames and words are compared with the oracle. As no filter has been here applied with respect to the corpus, all the 527 < LU, Frame > pairs in the manually created oracle have been used\(^6\), although only 437 pairs were represented through the MultiWordNet resources. The system was not able to decide for 71 words, and produced wrong guesses for 49 words: 317 correct guesses are thus produced. The results is a Precision of 0.87 and a Recall of 0.72. The recall value, lower than the coverage observed in the previous test, is also by a significant generative effect: the method discovers a number of new entries not accounted for in the oracle.

Typical new LUs introduced in the oracle are words not accounted for in the English Frameset, as reported in Table 6. The table shows that most of the new guesses of the system are indeed highly plausible. They represent widely used dialectal forms (e.g. masseria in the frame BUILDING), jargon (e.g. duepezzi in CLOTHING), technical terms (e.g. polizza in TEXT) and specific nouns (e.g. autorimessa in BUILDINGS). Although it is certainly questionable if words like articolo_di_fondo (i.e. main article in a newspaper) are worth to be considering as LUs for the frame TEXT, it is clear that if the application does main requires frame-like information, the presented model (even only the paradigmatic association here discussed) provides an effective tool for fast and robust prototyping. Notice that the low Recall (only 0.60 of the oracle words are correctly addressed), can be compensated by combining paradigmatic and distributional similarity (Equation 6), as experiments reported in Figure 3 suggest. We leave this last point as a future work.

4 Conclusions

In this paper we presented a combined model for representing frame semantics through paradigmatic and distributional evidence. We reported three experiments, which indicate possible application scenarios of these models. First, the combination of the presented models has been applied to extend FrameNet in a LU induction task, for English and Italian. In both cases the evaluation has shown that the combination of the two models achieves better performance against their independent uses, and that the level of accuracy is high enough to support lexicographers in the task of building FrameNets. In a second experiment, we showed that a strong association exists between lexical senses, as defined by WordNet, and the frame’s lexical units in FrameNet. Its automatic detection, as proposed in this paper, results in a significant reduction of the polysemy of LUs and in a highly accurate selection of those lexical senses semantically related to the situations represented by a frame. Finally, we demonstrated that this paradigmatic information can be used to develop a FrameNet resource in another language. For Italian, we automatically generated a very large and accurate set of 15,000 LUs.

The overall framework has encouraged us to develop a robust toolbox for the large scale acquisition of FrameNet-like lexicons in different domains and languages. The tool will be made publicly available for research studies in this area. Future work is on going on the adoption of richer models for Frameset, able to take into account more evidence than LUs, such as frame elements and syntagmatic information.
Moreover, the use of the derived space as a model for the recognition of frames in free-texts is expected to speed-up the development of a large collection of annotated sentences for the Italian language.

References

Agirre, E. and G. Rigau. 1996. Word sense disambiguation using conceptual density. In Proceedings of COLING-96, Copenhagen, Denmark.

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In Proceedings of COLING-ACL, Montreal, Canada.

Basili, R., M. Cammisa, and F.M. Zanzotto. 2004. A semantic similarity measure for unsupervised semantic disambiguation. In Proceedings of LREC-04, Lisbon, Portugal.

Basili, Roberto, Diego De Cao, Paolo Marocco, and Marco Pennacchiotti. 2007. Learning selectional preferences for entailment or paraphrasing rules. In Proceedings of RANLP 07.

Burchardt, Aljoscha, Katrin Erk, and Anette Frank. 2005. A wordnet detour to framenet. In Proceedings of the GLDV 2005 GermaNet II Workshop, Bonn, Germany.

Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Pad, and Manfred Pinkal. 2006. The salsa corpus: a german corpus resource for lexical semantics. In Proceedings of the 5th International Conference on Language Resources and Evaluation, Genova, Italy.

Burchardt, Aljoscha, Marco Pennacchiotti, Stefan Thater, and Manfred Pinkal. 2008. Assessing the impact of frame semantics on textual entailment. Journal of Natural Language Engineering (to appear).

Clark, Peter, Phil Harrison, John Thompson, William Murray, Jerry Hobbs, and Christiane Fellbaum. 2007. On the role of lexical and world knowledge in rte3. In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pages 54–59, Prague, June. Association for Computational Linguistics.

Harris, Zellig. 1964. Distributional structure. In Katz, Jerrold J. and Jerry A. Fodor, editors, The Philosophy of Linguistics, New York. Oxford University Press.

Heyer, L.J., S. Kruglyak, and S. Yooseph. 1999. Exploring expression data: Identification and analysis of coexpressed genes. Genome Research, (9):1106–1115.

Koehn, P. 2002. Europarl: A multilingual corpus for evaluation of machine translation. Draft.

Landauer, Tom and Sue Dumais. 1997. A solution to plato’s problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological Review, 104:211–240.

Lin, Dekang and Patrick Pantel. 2001. DIRT-discovery of inference rules from text. In Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD-01), San Francisco, CA.

Miller, G., R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. 1990. An on-line lexical database. International Journal of Lexicography, 13(4):235–312.

Pado, Sebastian and Mirella Lapata. 2007. Dependency-based construction of semantic space models. Computational Linguistics, 33(2).