Mapping CPA Patterns onto OntoNotes Senses

Octavian Popescu, Martha Palmer, Patrick Hanks

FBK-Irst, University of Colorado, University of Wolverhampton

popescu@fbk.eu, Martha.Palmer@colorado.edu, patrick.w.hanks@gmail.com

Abstract

In this paper we present an alignment experiment between patterns of verb use discovered by Corpus Pattern Analysis (CPA; Hanks 2004, 2008, 2013) and verb senses in OntoNotes (ON; Hovy et al. 2006, Weischedel et al. 2011). We present a probabilistic approach for mapping one resource into the other. Firstly we introduce a basic model, based on conditional probabilities, which determines for any given sentence the best CPA pattern match. On the basis of this model, we propose a joint source channel model (JSCM) that computes the probability of compatibility of semantic types between a verb phrase and a pattern, irrespective of whether the verb phrase is a norm or an exploitation. We evaluate the accuracy of the proposed mapping using cluster similarity metrics based on entropy.

1. Introduction

Most state-of-the-art systems in NLP now rely on annotated corpora. For certain NLP tasks such as PoS tagging, statistical machine-learning algorithms attain human-comparable accuracy. However, real-world NLP applications such as message understanding and high-quality machine translation require more complex text analysis, to facilitate the extraction and merging of various types of information. In order to manage this complexity and avoid the pitfalls of overtraining and/or overgeneralization, a system must be able to learn from rich and complementary annotation schemas (Bair et al. 2002). That is one reason why there has been a constant interest in the computational semantics community in the alignment of different resources (see, among others, Sinha & Mihalcea 2001, Palmer 2007, Baker & Fellbaum 2009, Wu & Palmer 2011).

In this paper we present an alignment experiment between corpus pattern analysis, CPA (Hanks 2005, Hanks 2013), and OntoNotes, ON, verb senses (Hovy et al. 2006, Weischedel et al. 2011). CPA has shown that, while the number of possible uses of each word may in principle be limitless, the number of normal uses is comparatively small and manageable. In other words, most lexical items habitually participate in only a comparatively small number of patterns. (A pattern corresponds to a prototypical usage: a combination of valency and collocation.) CPA is built from the bottom up by analysing and classifying, for each verb, a sample of several hundred actual uses of each verb in a corpus (BNC). The lexicographer´task is to find an appropriate level of generalization for each pattern and its implicatures (i.e., its meaning or entailments), while selecting an appropriate semantic type as an address or governing mechanism for clusters of collocates (lexical sets) used in the description of each pattern. However, it should be noted that actual verb usage in text varies on a continuum between regular constructs with clearly distinct senses at one end and highly innovative usages at the other extreme. The annotated examples in any CPA output are divided between norms and exploitations. The exploitation category contains examples in which the normal pattern has been "stretched" in order to accommodate writers´ and speakers´ creativity. The dividing line between norms and exploitations is fuzzy and therefore in a few cases the distinction is arbitrary.

In OntoNotes the WordNet senses of verbs have been clustered together in order to increase their utility in NLP tasks (see Section 2). By automatically annotating the ON sentences with CPA patterns, we achieve both a mapping between CPA and ON verb senses and a description of verb arguments using semantic features. Thus we have the benefit of obtaining a large resource where patternable word usage and shallow syntactic-semantic structures are linked together via grouped WordNet senses. By aligning these two independently developed systems of verb use and verb meaning, we aim to create a useful resource for machine-learning algorithms applied to meaning-related tasks.

There are two main issues that need to be addressed in aligning OntoNotes and CPA. Firstly, there is no pre-existing dictionary of lexical items that maps them onto CPA´semantic types. Thus, for any given word, there is no prior way to tell (1) what semantic type(s) it may have and (2) in which prototypical pattern(s) it is likely to appear. Such a dictionary is being created as part of the mapping process. Secondly, there is no relationship between ON senses and the semantic types of their arguments. This relationship must be learned. The fact that in OntoNotes the WordNet senses are grouped is instrumental for this goal. SUMO (Niles & Pease 2001) and PropBank (Palmer et al. 2005) are used in order to address these issues.

We present a probabilistic approach for resolving the two issues just mentioned. Firstly we introduce a basic model, based on conditional probabilities, which determines for a given sentence what is the best CPA pattern match. On the basis of this model, we propose a joint source channel model (JSCM), which computes the probability of compat-

ibility of semantic types between a verb phrase and a pattern, irrespective of whether the verb phrase is a norm or an exploitation. The JSCM model is not a mapping from exploitation to norm, but rather it is a way to generate and account for both the norm and the exploitation under the same constraints. These constraints describe compatible and potentially compatible combinations between the semantic types of arguments as a probability distribution.

The remainder of this paper is organized as follows: in the next two sections we present the two resources (OntoNotes and CPA). In Section 4 we present the statistical model used to realize the CPA-to-VerbNet mapping and evaluate the accuracy of this mapping. The paper ends with a section on conclusions and possibilities for further research.

2. OntoNotes

The OntoNotes (ON) project addressed the challenge of large-scale, accurate, and integrated annotation of shallow semantic structure in text. Funded by DARPA GALE and developed over several years, the OntoNotes project includes layers of annotation for nominal entity tags, syntactic structure, semantic role labeling, word sense, and coreference (Weischedel et al. 2011) and is available from the Linguistic Data Consortium. In this project we primarily make use of sense tags that are based on groupings of WordNet senses (Miller 1995; Fellbaum 1998). For the purpose of annotating various text genres, WordNet's fine-grained sense distinctions are not amenable to high rates of agreement among human annotators or to high automatic tagging performance (Palmer et al. 2007), which suggests that for many applications a coarser sense inventory is desirable.

OntoNotes represents an effort to create verb sense distinctions at a middle level of granularity that allow one to capture as much information as possible re a lexical item while still attaining high inter-annotator agreement (IAA) scores and high system performance in automatic sense disambiguation (Dligach & Palmer 2008). To date, ON has grouped over 2,500 English verbs, corresponding to most of the WordNet verbs that have three or more senses. Over 150,000 corresponding tokens were sense-tagged, with an average IAA of 89%. ON has demonstrated that coarse, clear sense distinctions can improve annotator productivity and accuracy. The performance of automatic systems typically lags some 10% behind IAA rates; human IAA scores of at least 90% for a majority of our sense-groupings resulted in the expected corresponding improvement in system performance, now well over 85% (Dligach & Palmer 2011), approaching human performance. (Brown et al. 2010) confirmed that the improvement is not simply based on having fewer classes to choose from.

Although mappings to WN are usually one-to-many, mappings have also been made to PropBank (Palmer et al. 2005), FrameNet (Fontenelle 2003), and VerbNet (Kipper et al. 2006), and these are often one-to-one. These mappings can be accessed via the SemLink database at the University of Colorado (Brown & Palmer 2010; Loper et al. 2007). SemLink currently includes the approximately 2,500 verbs linked to the verbs in PB. The coarse-grained OntoNotes senses have been mapped, where appropriate, to lexical units in FN frames and to VN classes. SemLink can be viewed through the Unified Verb Index website at http://verbs.colorado.edu/verb-index/index.php. Currently, 5,879 verbs (many that are monosemous) are represented in the index, which includes 5,726 links to VerbNet, 4,592 to PropBank, and 4,186 to FrameNet.

3. CPA

Corpus Pattern Analysis (CPA) is a technique for linking word use to prototypical syntagmatic patterns. Meanings and entailments are associated with patterns rather than with words in isolation. The technique is corpus-driven and is theoretically supported by the Theory of Norms and Exploitations (Hanks 2004, 2008, 2013). Pilot projects (2006-2010) were supported by the Czech research funding agency (GACR); current work in the UK is supported by AHRC. In the PDEV resource[1], created by CPA, verb patterns consist not only of basic argument structures (or valencies), but also of collocations and subvalency features. Collocates in each argument are grouped into lexical sets according to their semantic type. For some verbs, the semantic type of an argument is the only way to tell the difference between two senses of a verb. For example, *executing the prisoners* and *executing an order* are both transitive uses of the verb *execute*, but they have very different meanings. In the first example, the semantic type is [[Human]], while in the second it is [[Command]].

A shallow ontology of semantic types organizes them hierarchically. Each semantic type in the shallow ontology[2] is populated by a set of nouns derived from corpus analysis (rather than from introspective speculation). The semantic types of the arguments in each clause role collectively disambiguate each verb. To exemplify, consider the following BNC sentences containing the verb *accommodate*:

I'm sure he'll accommodate you

the attic is too large to accommodate the inscription.

pre-1979 Conservatives, by accommodating themselves to collectivist reforms

The following patterns match these sentences unambiguously:

[[Human 1]] accommodate [[Human 2]]

[[Building]] accommodate [[Physical Object]]

[[Human]] accommodate [[Self]] to [[Eventuality]]

| Patterns	Prototypical examples
[[Human]] | [[Institution]] abandon [[Activity]] | [[Plan]]	are incapacitated or have abandoned their practices he abandoned plans of working We should not abandon the search
[[Human]] | [[Institution]] abandon [[Attitude]]	he had abandoned immediate hopes abandoned their principles he had abandoned his commitment to persuasion
[[Human]] | [[Group]] abandon [[Location]]	citizens of Phocaea abandoned their town The lands that they abandoned before abandoning the site

Table 1: Patterns and Prototypical Examples

The associated Theory of Norms and Exploitations (TNE) says that language is indeed a form of rule-governed behavior, but argues that there is not just one monolithic system of rules. Instead, there are two interactive sets of rules: 1) Norms: a set of rules for using words normally and idiomatically: these are the rules of grammar augmented by rules governing normal collocation; they can account for about 90% of all utterances, but they do not account for linguistic creativity; 2) Exploitation rules, which account for creative and innovative usage (about 10% of everyday usage in a large general corpus such as BNC). Interaction over time between rules governing norms and rules governing exploitations of words accounts for a great deal of meaning change and resultant polysemy.

In CPA patterns are semantically motivated and are associated with prototypical sentence contexts (Hanks & Pustejovsky 2004, 2005). By analysing a sample of several hundred verb uses in the British National Corpus (BNC), sets of patterns have been extracted for each verb[3] In Table 1 the first column presents three patterns of the verb *abandon* and in the second column we show prototypical examples. $[[Human]]$, $[[Institution]]$, $[[Group]]$, etc., are semantic types, which represent features that each argument must have in order for the pattern to match the verb phrase. There are currently 226 semantic types used in CPA, organized in a shallow ontology (CPASO). These have proved sufficient to disambiguate all senses of all the 1,000 verbs which have so far been analyzed. The number of semantic types in this shallow ontology grows slowly as research proceeds. Every now and again a new semantic type must be added. It seems likely that, when all verbs have been analysed, an ontology of around 250 semantic types will be sufficient for disambiguation.

A semantic type outside a pattern is not functional but a word may have more than one semantic type, so recognizing the right pattern of a verb used in text necessitates recognizing the right semantic type. The examples in the second column in Table 1 represent an exact match of the patterns; they are norms (normal usage according to those

Figure 1: Distribution of number of patterns
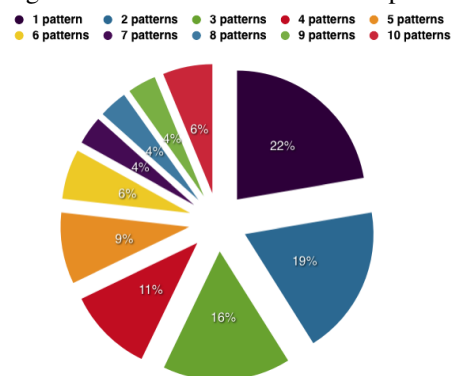
patterns). However, for an example like Ex1, which is considered an exploitation of the pattern on the first row, the matching is not exact. The semantic types associated with the word *path* are not [[Human]], [[Plan]], or any of the semantic types occurring in norm examples with this verb.

Ex1... movement towards abandoning the well-known paths

The CPA resource is growing continuously, the long-term goal being to achieve coverage of all the English verbs that are in normal use. There are approximately 6,000 such verbs in English, including predicative adjectives such as *be afraid* and *be glad*. At the time of writing, patterns for just over 1000 English verbs are available on the PDEV web site. The number of patterns for each verb varies between one for simple verbs such as *sentence* and over two hundred for light verbs such as *take*. Light verbs have not yet been fully analysed.

Figure 1 shows the distributions of the number of patterns in CPA. Two of the semantic types in CPASO, [[Human]] and [[Institution]], are significantly more frequent than others; they are used 1,849 and 365 times respectively. PDEV also provides information about the comparative frequency of the patterns of each verb. The distribution of the patterns in corpus is not uniform, the mode being that a dominant pattern is likely to have many more occurrences than the next most frequent pattern.

We computed how many times the dominant pattern for a

---

[3]CPA is currently being extended into the analysis of words and meanings in several languages, including Italian and Spanish.

verb has more than 40%, 60% or 80% of the occurrences in the sample, by also considering the total number of patterns for the respective verbs grouped in intervals: verbs that have between 3 and 5 patterns, verbs that have between 6 and 20 patterns, verbs having between 21 and 40 patterns, and verbs having between 41 and 60 patterns. For example, 65.25% of the verbs with between 6 and 20 patterns have a dominant pattern that occurs more than 40% in the corpus, but only 23.72% of the verbs with the same number of patterns have a dominant pattern that occurs more than 60% of the time in the corpus. See Table 2.

We decided to explore mapping from SUMO to the semantic types in the CPA Shallow Ontology (CPASO), as proposed in (Popescu 2013, Kawara et al. 2014). The SUMO ontology is aligned to the senses of Wordnet1.6. In Table 3 we list the SUMO attributes for the direct object of the examples in Table 1.

The mapping from CPA to SUMO is one to many. The pattern learning and recognizing algorithm must be able to retain for a word only the SUMO attributes which are instantiated in a particular example. The algorithm presented in the next section learns which SUMO attributes are relevant for a CPA pattern for each word.

4. Mapping CPA to ON

In this section we present two probabilistic models for mapping CPA to ON. The first model considers only the norm patterns and it computes the conditional probability for a given phrase to be matched by a certain pattern given the CPA examples and the mapping to SUMO. This mode is considered to be the basic model. The second model extends the basic models considering also the exploitation examples and it is built in the joint source channel framework. In Subsection 4.1 the first model is explained and then, in the following subsection, the JSCM model is explained.

1st frequent	2-5	6-20	21-40
40%	94.35%	65.25%	25%
60%	60.45%	23.72%	12.5%
80%	27.1%	14.23%	0%

Table 2: Dominant Pattern Frequency in Corpus

direct object	SUMO attributes
plan	Plan, Abstract, icon
practice	NormativeAttribute, EducationalProcess
search	Pursuing, Investigating
hope	EmotionalState, Reasoning
principle	NormativeAttribute, Proposition
commitment	TraitAttribute, Declaring
town	City, Geopolitical
land	LandArea, Geopolitical, Nation
site	LandArea, Located

Table 3: Mapping CPA semantic types to SUMO

4.1. Conditional Probability Model

Matching a corpus pattern against a verbal phrase involves labeling the heads of the constituents with semantic features and the verb with a pattern number. We build a probabilistic model in which we compute the probability in Equation (1),

$$p(t_0, t_1, t_2, t_3, .., t_n, w_1, w_2, w_n) \qquad (1)$$

where t_0 is the pattern number, t_i is the semantic type of the word w_i, which is the head of the ith constituent, with i from 1 to n. For a given sentence we choose the most probable labelling, Equation (2)

$$p(t_0^c, t_1^c, t_2^c, t_3^c, .., t_n^c, w_1^c, w_2^c, w_n^c) = \arg\max_{t_i} p(t_0, \text{ w}_n) \qquad (2)$$

Following the technique described in Popescu (2007, 2013), we build a mapping from CPA semantic types to SUMO attributes. In this way we can associate a set of semantic types with any English word (other than function words). First, we construct the confusion matrix for each slot, with SUMO attributes, and CPA patterns. Second, on the basis of the relationship existing between the senses of the fillers of the corpus pattern, called the Chain Clarifying Relationship (CCR), and the fact that the patterns have a regular language structure, we learn for each verb its discriminative patterns with SUMO attributes. Using the chain formula, and grouping the terms conveniently, Equation (1) becomes Equation (3).

$$p(t_0, t_1, t_2, t_3, .., t_n, w_1, w_2, w_n) = p(t_0)p(w_1|t_0)...$$
$$... p(t_n|t_0, w_1, t_1, w_2, ..., t_{n-1}, w_n)$$

$$\simeq p(t_0)p(w_1|t_0)p(t_1|t_0, w_1)p(w_2|t_0)p(t_2|t_0, w_2)...$$
$$... p(t_n|t_0, w_n)$$

$$\simeq p(t_0)p(w_1|t_0)p(t_1|t_0)p(t_1|w_1)p(w_2|t_0)p(t_2|t_0)p(t_2|w_2) ...$$
$$... p(t_n|t_0)p(t_n|t_0)p(t_n|w_n)$$

$$(3)$$

The quantities on the right-hand side are computed as follows:

- $p(t_0)$ is the probability of a certain pattern. This probability is given in CPA

- $p(w_i|t_0)$ is the probability that a certain word will be the head of a constituent. It is computed from the examples associated with each pattern with a smoothing technique similar to the one used in N-gram models (Cheng et al. 2004).

- $p(t_i|t_0, w_i)$ is the probability that a certain word at a certain position in the pattern will carry a specific semantic type. This probability is equated to $p(t_i|w_i)p(t_i|t_0)$, assuming independence between the verb sense and the word given the semantic type. The

885

first of the two later probabilities is extracted from Semcor (Miller et al. 1993), while the second comes from the confusion matrix mentioned above.

4.2. Joint Source Channel Model

The verbal phrases can be segmented naturally into segments, each segment representing a grammatical function together with a semantic type. From this point of view, a pattern is a VP for which the segments are known. An arbitrary VP may admit different segmentations, out of which only one is the correct one. The problem of recognizing the pattern that matches a verbal phrase is a problem of finding the most probable alignment between the segments associated with the verbal phrase and the corpus pattern, PT. The probability $P(VP, PT)$, is computed using Equation (4).

$SS \quad possible\ segments\ for\ VP$
$PP \quad segments\ of\ a\ given\ pattern\ PT$

$$
\begin{aligned}
P(VP, PT) &= P(ss_1, ..., ss_n, pp_1..., pp_n) \\
&= P((ss, pp)_1, ..., (ss, pp)_n) \\
&= \prod_{i=1}^{n} P((ss, pp)_i | (ss, pp)_1^i)
\end{aligned}
\tag{4}
$$

Equation (4) can be marginalized in terms of both SS and PP, therefore obtaining both the probability that a pattern matches the VP and the back probability that the VP leads to a certain pattern. In future work we can use Equation (4) to induce patterns in an unsupervised way and/or in a bilingual framework considering CPA for two languages. In this work, we consider the patterns as given, they are the ones created in CPA, thus PP is constant over all alignments. Consequently, and we search for SS and γ that maximize (5), where γ is an alignment from SS to PP:

$$
SS_{OT} = \underset{SS, \gamma}{\arg\max}(VP, PP, SS, \gamma)
\tag{5}
$$

When the same segments are present both in the VP and in the pattern the alignment is trivial and in this case we have an example of norm usage. However, the segments can be different and in this case the alignment is a probability distribution over the possible choices for that particular argument of the VP function. The probability of segment alignments is learned from the examples provided by CPA. As the examples are from raw text, the probability that a given word will have a given semantic type must be computed. Such probabilities are computed via the mapping of the CPA semantic types to SUMO, see Section 4.1. The probability of a certain word, W, occurring as an argument, ST, of a VP that is matched by a certain pattern, PT, is computed using Equation (6):

$$
P(PT, ST, w) = P(PT)P(w|PT)P(ST|w, PT)
\tag{6}
$$

In practice we use the equation after parsing the text, and therefore the $P(w|PT)$ is a characteristic function, P = 0 or 1 according to the parse tree. We also use the semantic independence assumption between the probability that a certain word has a certain SUMO feature and the probability that that SUMO feature is used in the given pattern. Equation (6) is rewritten as Equation (7):

$$
P(PT, ST, w) \simeq P(PT)P(ST|w)P(ST|PT)
\tag{7}
$$

The quantities on the right-hand side of Equation 7 can be computed from an annotated corpus. The probability $P(ST|w)$ can be computed from Semcor (Moldovan 1999). The probability $P(ST|PP)$ is given in CPA as estimated from sampling from BNC.

Training the system is a two-step process:

(S1) Equation (7) is used only on the norm CPA examples, thus the conditional probabilities of patterns given the SUMO semantic types are computed.

(S2) Equation (4) is used to align the semantic types not presented explicitly in the patterns to those present in the patterns using both norm and exploitation CPA examples.

The output of the training process is a set of SUMO attribute pairs which are ordered according to their probability. These pairs represent the constraints for the relevant verb. Some of these constraints are global, like the constraint that a SUMO attribute can be replaced with its hyponym, but they can be overridden by local constraints for particular verbs.

We rely on cluster evaluation methods in order to test the JSCM approach trained on CPA against the OntoNotes senses (Li et al. 2004). We looked at E - entropy, RC - recovery rate, and P - purity. The consistency of pattern recognition inside each cluster of examples annotated with the same OntoNotes represents a good evaluation of the capacity of the approach to predict verb usage.

$$
e_j = \sum_{i=1}^{L} p_{ij} \log_2 p_{ij}, p_{ij} = \frac{m_{ij}}{m_i}
\tag{8}
$$

$$
E = \sum_{j=1}^{K} \frac{m_j}{m} \cdot e_j, RC = 1 - \sum_{j,i=1}^{K,L} \frac{p_{ij}}{m_i}
\tag{9}
$$

$$
P = \sum_{j=1}^{K} \frac{m_j}{m} \cdot p_j, p_j = \max_i p_{ij}
\tag{10}
$$

where m_j is the number of elements in cluster of OntoNotes verb sense j, m_{ij} is the number of elements in cluster j and class of CPA pattern i, m is the total number of elements, L is the number of recognized CPA patterns, K is the number of OntoNotes senses. Table 4 presents the macro-average (Tsoumakas et al. 2010). We preferred here to use the macro-average for counterbalancing the effect of masking the less frequent CPA patterns (Manning et al. 2008).

In order to understand better the relationship between the number of training examples and the accuracy the result,

	measure	value
[Conditional]	Entropy	0.405
	Recovery Rate	0.866
	Purity	0.77

	measure	value
[JSCM]	Entropy	0.427
	Recovery Rate	0.788
	Purity	0.805

Table 4: CPA vs. ON cluster validation

Figure 2: Variation of Entropy
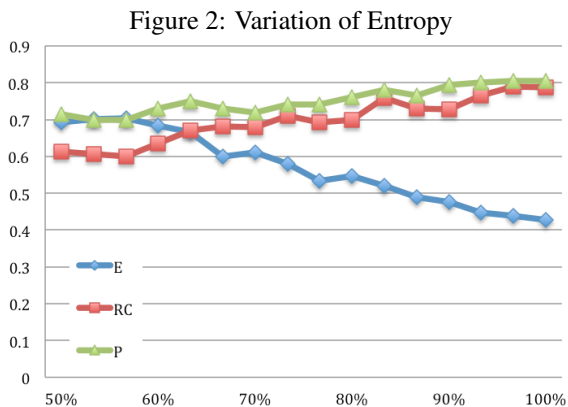
we calculated the *E*, *RC* and *P* for various coverage of the CPA examples in training. The plots in Figure 2 show that there is not a linear relationship between the number of examples and the quality of clustering. Some CPA examples do not match the OntoNotes senses, so it would be better to isolate those examples. This may be due to different views on senses, or possibly to misclassified examples in corpora. More research is needed to clarify this issue; we plan to investigate it in future research.

5. Related Work

There is a vast literature on semantic frames, and unfortunately we cannot exhaustively refer here to all the important papers. The CPA resource is hand-crafted, as are semantic frames in the lexicons of FrameNet (Baker et al. 1998) and PropBank (Palmer et al. 2005). However, there are notable differences between the latter ones and CPA. CPA has been created from the bottom up, starting from a large sample of corpus examples. There was no pressure to align the patterns to a prior given collection of semantic roles or frames. Our task of pattern recognition in verb usage can be regarded as based on clustering of verb instances. Parisien and Stevenson (2009) proposed a Dirichlet Process model for clustering usages of the verb *get*. Later, Parisien and Stevenson (2010) proposed a Hierarchical Dirichlet Process model for jointly clustering verb arguments into classes. However, their argument structures are not semantic but syntactic, and also they did not evaluate the resulting frames.

There have also been related approaches to clustering verb types (Vlachos et al. 2009; Sun and Korhonen 2009; Falk et al. 2012; Reichart and Korhonen 2013). These methods induce verb clusters in which multiple verbs participate, and do not consider the polysemy of verbs. Another line of

related work is unsupervised semantic parsing or semantic role labeling (Poon and Domingos 2009; Lang and Lapata 2010; Lang and Lapata 2011a; Lang and Lapata 2011b; Titov and Klementiev 2011; Titov and Klementiev 2012). These approaches basically cluster predicates and their arguments to distinguish predicate senses and semantic roles of arguments. However, they did not aim at recognizing similar usage patterns, but at distinguishing verbs that have different senses in a relatively small annotated corpus. Applying this method to a large corpus could produce a frame lexicon that is too big to be useful. Also, the scalability of their method may be an issue. The main focus of our research in this paper is the difference between norm and exploitation and to build a model able to cope simultaneously with the two categories of verbal usage under the same assumptions. The results obtained strongly suggested that this distinction is important and simply considering a training corpus of mixed examples is likely to produce a poorer model than the JSCM model presented here.

Zhao, Meyers, and Grishman (2004) proposed a SVM application for slot detection, which combines two different kernels, one of them being defined on dependency trees. Their method tries to identify the possible fillers for an event, but it does not attempt to treat ambiguous cases; also, the matching score algorithm makes no distinction between the importance of the words, considering equal matching score for any word within two levels of the dependency tree. Many of the purely syntactic methods have considered the properties of the subcategorization frames of verbs. Verbs have been partitioned in semantic classes mainly on the basis of Levin classes and their alternations (Dorr & Jones 1996, Dang et al. 1998, Collins 1999, McCarthy 2001, Korhonen 2002, Lapata & Brew 2004). These semantic classes can be used in WSD via a process of alignment with hierarchies of concepts as defined in sense repository resources (Shin & Mihalcea 2005). However, the problem of the consistency of alignment is still an open issue.

6. Conclusion and Further Work

In this paper we have presented work carried out on the alignment of two resources that describe the behavior of verb phrases in natural language. By mapping CPA to ON senses, we linked prototypical patterns of verb usage to classes of verbs created by introspection. On the basis of previous work undertaken in the OntoNotes project, the relationship between semantic types, corpus examples for each pattern, and WordNet and FrameNet may be studied further. In this way, the machine-learning algorithms ben-

efit from a resource that is instrumental in the computation and management of various aspects of meaning, in order to increase the accuracy of semantic tasks. In the near future we would like to improve the accuracy of the mapping. We will focus on two aspects: (1) improving the statistical model by considering methods that can constrain the confusion matrix, such as Maximum Entropy and/or Minimum Description Length, and (2) automatically constructing a set of semantic types by clustering into topics syntactic frames extracted automatically from corpora, following Materna (2012).

One goal that we consider important is to extend the JSCM model to account for new patterns by implementing a bootstrapping approach. Learning the most probable set of constraints solely on the raw text occurrence may be carried out by employing unsupervised clustering using EM algorithm. There is a direct link between the JSCM mapping and the semantic constraints used in optimality theory. On the basis of this connection we could develop a pattern matching algorithm. Another direction of research is to check the consistency and the soundness of the found constraints against a corpus annotated with semantic text similarity scores.

7. Acknowledgments

8. References

Eric Bair, Trevor Hastie, Debashis Paul, Rob Tibshirani 2006. *Prediction by supervised principal components.* In J. Amer. Statist. Assoc. (2006), vol. 101

Collin F. Baker, Christiane Fellbaum 2009. *WordNet and FrameNet as Complementary Resources for Annotation.* Linguistic Annotation Workshop 2009

Stephan Bloehdorn, Alessandro Moschitti 2006. *Combined Syntactic and Semantic Kernels for Text Classification.* Advances in Information Retrieval, Springer

M. Collins 1999. *Head-Driven Statistical Models for Natural Language Parsing.* Ph.D. thesis, University of Pennsylvania

T. Dang, K. Kipper, K. Palmer, J. Rosenzweig 1998. *Investigating regular sense extensions based on intersective Levin classes.* Coling-ACL98, Montreal, August 11-17

Dmitriy Dligach, Martha Palmer: 2011. *Good Seed Makes a Good Crop: Accelerating Active Learning Using Language Modeling.* In Proceedings of ACL

B. Dorr, D. Jones 1999. *Acquisition of Semantic Lexicons* In Breadth and Depth of Semantic Lexicons, Edited by Evelyne Viegas. Kluwer

Ingrid Falk, Claire Gardent, Jean-Charles Lamirel 2012. *Classifying French verbs using French and English lexical resources.* In Proceedings of ACL2012, pages 854-863

Thierry Fontenelle, ed. 2003. *Special Issue on FrameNet.* International Journal of Lexicography, volume 16.

Patrick Hanks 2004. *Corpus Pattern Analysis.* In Proceedings of Euralex 2004.

Patrick Hanks 2008. *Mapping Meaning onto Use: a Pattern Dictionary of English Verbs.* In Proceedings of AACL 2008.

Patrick Hanks 2013. *Lexical Analysis: Norms and Exploitations.* MIT Press 2013

Patrick Hanks, James Pustejovsky 2004. *Common Sense about Word Meaning: Sense in Context.* In Proceedings of TSD 2004

Patrick Hanks, James Pustejovsky 2005. *A Pattern Dictionary for Natural Language Processing.* In Revue Franaise de Langue Applique 10:2.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, Ralph Weischedel 2006. *OntoNotes: The 90% Solution.* In Proceedings of HLT-NAACL

D. Kawara, D. Pederson, O. Popescu, and M. Palmer 2014. *Inducing Example-based Semantic Frames from a Massive Amount of Verb Uses.* In Proceedings of EACL

Karin Kipper, Anna Korhonen, Neville Ryant, Martha Palmer 2006. *A Large-scale Classification of English Verbs.* Language Resources and Evaluation Journal, 42(1). Springer Netherland

A. Korhonen 2002. *Subcategorization Acquisition.* PhD thesis, published as Technical Report UCAM-CL-TR-530. Computer Laboratory, Cambridge

Joel Lang, Mirella Lapata 2011b. *Unsupervised semantic role induction with graph partitioning.* In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing

M. Lapata, C. Brew 2004. *Verb Class Disambiguation Using Informative Priors.* Computational Linguistics 30:1, 45-73

Yoong Keok Lee, Hwee Tou Ng, Tee Kiah Chia 2004. *Supervised Word Sense Disambiguation with Support Vector Machines and Multiple Knowledge Sources.* In Proceedings of SenseEval-3.

Tao Li, Sheng Ma, Mitsunori Ogihara 2004. *Entropy-Based Criterion in Categorical Clustering.* In Proceed-

ings Categorical Clustering of the 21st International Conference on Machine Learning, Banff, Canada

Dekang Lin 1998. *An information-theoretic definition of similarity.* In Proceedings of the Fifteenth International Conference on Machine Learning.

Edward Loper, Szu-ting Yi and Martha Palmer 2007. *Combining Lexical Resources: Mapping Between PropBank and VerbNet* In Proceedings of the 7th International Workshop on Computational Semantics, Tilburg

J. Materna 2013. *Parameter estimation for LDA-Frames.* In Proceedings of the Thirtieth International Conference on Machine Learning, Atlanta, GA. In Proceedings of NAACL-HLT2013.

D. McCarthy, J. Carroll, J. Preiss 2001. *Disambiguating noun and verb senses using automatically acquired selectional preferences.* In Proceedings of the SENSEVAL-2 Workshop, ACL/EACL, Toulouse, France

Hoifung Poon, Pedro Domingos 2009. *Unsupervised semantic parsing* In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP2009)

Octavian Popescu, Bernardo Magnini 2007. *Sense Discriminative Patterns for Word Sense Disambiguation.* In Proceedings of the SCAR Workshop, NODALIDA

Octavian Popescu 2013. *Learning Corpus Pattern with Finite State Automata.* In Proceedings of the ICSC 2013.

Allan Prince, Paul Smolensky 1999. *Optimality Theory Constraint Interaction in Generative Grammar.* MIT Press 2013

James Pustejovsky, Elisabetta Jezek 2009. *Semantic Coercion in Language: Beyond Distributional Analysis.* Italian Journal of Linguistics 2009

Roi Reichart, Anna Korhonen 2013. *Improved lexical acquisition through DPP-based verb clustering.* In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL2013)

Lin Sun, Anna Korhonen 2009. *Improving verb clustering with automatically acquired selectional preferences.* In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP2009)

Ivan Titov, Alexandre Klementiev 2011. *A Bayesian model for unsupervised semantic parsing.* In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies

Andreas Vlachos, Anna Korhonen, Zoubin Ghahramani 2009. *Unsupervised and constrained Dirichlet process mixture models for verb clustering.* In Proceedings of the Workshop on Geometrical Models of Natural Language Semantics

Ralph Weischedel, Eduard Hovy, Mitchell Marcus, Martha Palmer, Robert Belvin, Sameer Pradan, Lance Ramshaw, Nianwen Xue 2011. *OntoNotes: A Large Training Corpus for Enhanced Processing.* Data Acquisition and Linguistic Resources of the Handbook of Natural Language Processing and Machine Translation: Global Automatic Language Exploitation

Shumin Wu, Martha Palmer 2011. *Semantic Mapping Using Automatic Word Alignment and Semantic Role Labeling.* In Proceedings of ACL11, Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation

S Zhao, A. Meyers, R. Grishman 2004. *Discriminative Slot Detection Using Kernel Methods.* In Proceedings of COLING-04