Advances in Cognitive Systems 5 (2017) Submitted 3/2017; published 5/2017

Kinesthetic Mind Reader: A Method to Identify Image Schemas in Natural Language

Dagmar Gromann DGROMANN@IIIA.CSIC.ES
Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Spain

Maria M. Hedblom MARIAMAGDALENA.HEDBLOM@UNIBZ.IT
Free University of Bozen-Bolzano, Bozen-Bolzano, Italy

Abstract

Natural language understanding remains one of the weak spots of Artificial Intelligence and cognitive systems in general. In cognitive linguistics, image schemas were introduced as spatio-temporal relations learned from sensorimotor processes that constitute conceptual building blocks for high-level cognition, such as language and reasoning. In this role, they have been successfully employed in computational concept invention and conceptual metaphor research. However, due to their abstract nature identifying them in natural language is an open challenge. To address this issue, this paper proposes a spectral clustering method combined with semantic role labeling to semi-automatically detect image schemas in natural language. In the majority of identified spatial clusters from the Europarl corpus the proposed method detected image schemas, which shows that it works effectively on large corpora. The outcome of this method is a repository of natural language expressions annotated with image schemas that can be used to improve spatial language understanding and provide examples to supervised machine learning approaches.

1. Introduction

Detecting spatio-temporal relations in natural language is central to a wide range of Artificial Intelligence (AI) applications, including manipulation instructions in human-robot interaction (Misra et al., 2016; Kollar et al., 2014), simulation of natural sensorimotor knowledge acquisition of infants (Guerin, 2008), and any kind of mapping between natural language symbols and their objects in the physical world (Krishnamurthy & Kollar, 2013). However, the symbol grounding problem of how signs obtain their meaning, relate to physical objects, and are cognitively represented, remains challenging. The cognitive grounding of spatial language has been investigated in different disciplines, including AI (Misra et al., 2016) and linguistics (Hampe & Grady, 2005). This paper contributes to the research on the grounding of spatio-temporal natural language by proposing a method to automatically detect cognitive conceptual building blocks in textual data.

Looking at formal representations of everyday concepts and events, the symbol grounding problem demonstrates some of the limitations of artificial systems. For instance, Morgenstern (2001) investigates the underlying meaning of the event ‘egg cracking’. Her formalization uses no less than 66 axioms to complete the representation. If this is the case for one simple event consider asking a formal system to conceptualize the event of 'making an omelet' or the more contextually complicated event of 'baking a wedding cake'. Reasoning can be done at different levels of abstractions and not each of these axiomatizations might need to be considered in the process. In contrast, embodied cognition views concepts as more compressed representation derived from the embodied experience of executing/perceiving such events. However, in order to achieve a formal representation of image schemas, their formal representation needs to be sufficiently fine-grained as well Hedblom et al. (2015).

One theory of such compressed representation of information goes under the name image schema (Johnson, 1987; Lakoff, 1987). Image schemas are spatio-temporal relations learned in early infancy that through analogical reasoning can be used to explain and predict outcomes on any current or future situation (Mandler & Pagán Cánovas, 2014). For instance, if a child has learned that it can go IN and OUT of a house – having learnt the image schema of CONTAINMENT – it can transfer this information to an 'egg cracking'-situation. Here the child can infer that the edible part of the 'egg' is contained within the shell and by breaking the border (as in 'open a door') the egg can get out. As a highly interdisciplinary research field, approaches to understanding image schemas are conducted in (cognitive) linguistics (e.g. Dodge & Lakoff (2005)), formal approaches (e.g. Bennett & Cialone (2014)) and developmental psychology (e.g. Mandler & Pagán Cánovas (2014)).

This paper rests on the assumption that any mental conceptual system guiding abstract thinking and acting in humans is also the system that communication is based on. We consider natural language "an important source of evidence of what that system is like" (Lakoff & Johnson, 1980), where that refers to the cognitive system guiding our actions. We use spectral clustering to separate spatial from non-spatial language with the assumption that spatial clusters potentially contain image schemas. To this end, we build on research from cognitive linguistics on spatial language (Talmy, 2005; Zlatev, 2010) that has found prepositions to be excellent spatial indicators. The proposed method clusters verb-preposition combinations with dependent nouns as features. Resulting clusters are successfully identified as spatial or non-spatial by means of semantic role labelling and manually analyzed to detect image schemas. As outcome this method provides a large repository of natural language expressions (verb-preposition-noun triples) annotated with image schemas. It is our belief that such a linguistically grounded yet formally specified repository of image schemas represents an excellent framework for cognitive knowledge acquisition and spatial language analyses.

2. Foundation and Related Work

Image schemas are defined as prelinguistic conceptual building blocks that allow higher cognition, such as language and reasoning, to be grounded in low-level sensations acquired from embodied experiences (Johnson, 1987). Some examples are: the UP-DOWN image schema, also called VERTICALITY; the notion of inside-border-outside, namely CONTAINMENT; the dependency relationship found in the image schema SUPPORT; and SOURCE_PATH_GOAL which capture movement between points. Image schemas are often described as spatio-temporal relations and were introduced in the area of cognitive linguistics focusing on spatial semantics (Zlatev, 2010). Talmy (2005) presents an influential perspective on how spatial schemas exist in language, focusing on spatial relations between two objects (or regions). An example is the sentence *"the lion* (called trajector) *ran after the deer* (called landmark)*"*. One of the most common word classes to indicate spatial schemas is prepositions. As illustrated in the example, the preposition 'after' demonstrates the spatial relationship between the lion and the deer. Ideally for a system that identifies spatial relations and dimensions in language, the prepositions could be mapped directly to particular spatial situations. Unfortunately, prepositions have been shown to be highly ambiguous and to realize multiple meanings in different contexts. Compare for instance the difference between the prepositional meaning in: *"the book is* on *the table"* (spatial) and *"a book* on *mircobiology"* (topic). The Preposition Project (TPP) (Litkowski & Hargraves, 2005) and the Pattern Dictionary of English Prepositions (PDEP) (Litkowski, 2014) identified 20 different categories of meaning for prepositions (e.g. spatial, temporal, topic and cause).

One method designed to extract spatial relations between the trajector and the landmark was introduced by Kordjamshidi et al. (2011). They suggest to use triples of words consisting of the 'the trajector', 'a spatial indicator' (the preposition) and 'the landmark'. Building on the TPP (Litkowski, 2014), their relations focus on the prepositions as the spatial grounding factor. While this method has a lot to offer it disregards the influence of verbs as moving indicators in their machine learning approach. This is problematic as verbs play a central role in identifying the relation between trajector and landmark (Kollar et al., 2014). Reconsider the example above where 'ran' provides crucial information of both temporal and spatial nature of the relationship between the lion and the deer.

Psychological support for image schemas comes from how they offer infants conceptual grounds to make predictions about their surroundings (Mandler & Pagán Cánovas, 2014; Gibbs & Colston, 1995). Indeed, work in linguistics (e.g. Dodge & Lakoff (2005)) and psychology (e.g. Mandler & Pagán Cánovas (2014)) reveal image-schematic involvement in reasoning and language development. In developmental psychology, the image schema demonstrates how key concepts are transferred through analogical reasoning and conceptual metaphors (Lakoff & Johnson, 1980). For example, if an infant has learned that 'tables SUPPORT plates', it can infer that 'desks SUPPORT books'. It is proposed that a similar method is applied when language is developed, in particular when abstract concepts are concerned. Statements such as *"to offer* SUPPORT *to a friend in need"* or *"to* put in *a good word"* provide some good examples. Pauwels (1995) went so far to claim that any abstract use of the word "put" requires the understanding of CONTAINMENT stressing the importance of verbs in image schema analyses.

Approaching image schemas from a formal perspective with the goal to integrate image schemas in computational systems dealing with natural languages, Hedblom et al. (2015) took a closer look at what in linguistic research is often referred to as the SOURCE_PATH_GOAL schema Lakoff & Núñez (2000). By looking at conceptual occurrences of movement and 'paths' the authors build a hierarchical ontology of what they introduced as the PATH-following family. The cognitive inspiration for the ontology was based on research in developmental psychology that show how image schemas are learned by adding spatial and temporal elements and their complexities through increased exposure to particular situations (Mandler & Pagán Cánovas, 2014).

Approaches for the extraction of image schemas from natural language generally focus on lexical surface structures. Working top-down Kuhn (2007) uses WordNet to extract image-schematic expressions from language and Dodge & Lakoff (2005) provide a detailed analysis of the image-schematic annotation of linguistic features across languages by using manually curated examples. Embodied Construction Grammar (ECG) (Feldman et al., 2009) combines the formalism of construction grammar with the cognitive theory of image schemas, however, exemplified rather than evaluated. Bennett & Cialone (2014) present a pattern-based extraction method based on synonyms of the image schema CONTAINMENT (e.g. 'surround', 'enclose'). Their method identified and formally represented eight different CONTAINMENT structures. A similar approach extracting SOURCE_PATH_GOAL structures from multilingual financial data was proposed by Gromann & Hedblom (2016). By extracting terms and their definitions the authors could extend the ontology on PATH-following previously introduced by Hedblom et al. (2015).

3. Methodology

To the best of our knowledge there exists so far no automated approach to image schema detection in natural language text. Following the state-of-the-art technique for exploratory data analysis (Baroni et al., 2014), we decided to opt for unsupervised clustering. The chosen normalized spectral clustering algorithm proposed by Ng et al. (2001) has been effectively applied to various lexical acquisition tasks (e.g. Shutova et al. (2016); Xu & Ke (2016); Sun & Korhonen (2009)). As detailed in Section 2, prepositions are major indicators for spatial schemas in natural language and verbs are the best approximation to detect movement. Nouns that co-occur with verb-preposition combinations in natural language text are used as features for the clustering algorithm, where the feature values are their relative frequencies. For instance, the feature vector for *bring-into* would be {*disrepute*: n_1, *play*: n_2, *force*: n_3, ..., *operation*: n_i} where n represents the frequencies. The verb-preposition combination as indicators for spatial schemas is also backed by manual corpus-based analyses of natural language from image-schematic research (e.g. Papafragou et al. (2006); Dodge & Lakoff (2005)). The resulting clusters are divided into spatial and non-spatial by using a semantic role labeling tool (Punyakanok et al., 2008). Spatial clusters are then analyzed manually for image-schematic structures.

3.1 Dataset

Our approach is based on the distributional hypothesis (Harris, 1954) which states that the co-occurrence of words in contexts indicates semantic similarity. This means that we require a large text collection to analyze a sufficient number of contexts for each verb-preposition combination. We used one such corpus, namely the Europarl corpus (Koehn, 2005). It is appropriate as it is commonly used, substantial in terms of size (1,959,830 sentences), and additionally offers a complete alignment across all the European languages of its text collection. The corpus contains sentences extracted from the proceedings of the European Parliament. Thus, the coverage of the corpus is somewhat limited to topics revolving around governance and political issues. While this might be a limited topical scope, we believe that the abstract character of the domain provides an excellent basis for the extraction of image schemas that are not obviously linked to the natural language expressions, such as *put on the table* which is clearly associated with SUPPORT as opposed to *put in a good word*. Additionally, as we are interested in differences and similarities of linguistic manifestations of image-schematic structures across languages, the corpus allows for an easy transition if further experiments were to be carried out cross-lingually.

3.2 Dependency Parsing

Building on dependency parsers and dependency relations in order to extract features for the clustering process is not new and has been successfully applied with verb-noun dependencies in conceptual metaphor extraction (Shutova et al., 2016). We perform dependency parsing on each sentence using the Stanford Dependency Parser (Chen & Manning, 2014). This process identifies how individual elements of a natural language sentence depend on each other and represents state-of-the-art preprocessing of natural language data to uncover spatial language (Kordjamshidi et al., 2011). We extract prepositions and the corresponding dependency relations to nouns, namely noun modifier relations (*nmod*), which in turn relate to the corresponding verb by means of nominal subject relations (*nsubj*, *nsubjpass*). When extracting verbs and nouns we analyze their direct dependencies in order to consider phrasal verbs and noun phrases, which is important for an accurate frequency analysis. For instance, *get* is not the same as *get up* and *union* is semantically distinct from *European Union*. We perform frequency analysis on all extracted triples of verb-preposition-noun combinations to build feature vectors for the similarity matrix.

3.3 Textual Semantic Distance

Computing a similarity matrix depends on the choice of semantic distance measure that is best for the given data. The most commonly used similarity measures between textual elements are Term Frequency-Inverse Document Frequency (TF-IDF), Positive Pointwise Mutual Information (PPMI), Kullback-Leibler divergence, string edit distances, and cosine distance. TF-IDF calculates the term frequency while considering its inverse document frequency, thereby calculating how important a term is for a given collection of text documents. Pointwise mutual information quantifies the difference between the probability of two textual units occurring together and the presumed co-occurrence under the independence condition. With x and y representing the frequencies of verb-preposition-noun combinations, we use Positive Pointwise Mutual Information (PPMI) as defined in Equation 1 to create a similarity matrix, where $PMI(x, y)$ is set to zero if its value is below zero.

$$PMI(x, y) = log \frac{p(x, y)}{p(x)p(y)} \tag{1}$$

The Kullback-Leibler divergence is a useful extension of PMI to measure mutual information that multiplies PMI with $p_{(x,y)}$. It has less bias towards rare-occurring phrases and is particularly adequate for the comparison of multi-word expressions. A symmetric and smoothed version of the Kullback-Leibler is the Jensen-Shannon divergence (JSD). Since the similarity matrix represents a weighted but undirected graph, the JSD is preferable to Kullback-Leibler and for the two feature vectors v_i and v_j is defined in Equation 2.

$$JSD(v_i||v_j) = \frac{1}{2}D(v_i||M) + \frac{1}{2}D(v_j||M) \tag{2}$$

Here D represents the Kullback-Leibler divergence and M is defined as the average of v_i and v_j. We adopt the successful creation of a similarity matrix by Shutova et al. (2016) and define the similarity w_{ij} in Equation 3:

$$w_{ij} = e^{-JSD(v_i,v_j)} \tag{3}$$

3.4 Clustering

Spectral clustering is particularly attractive since it is reasonably fast and transforms the data clustering into a graph partitioning problem. A similarity matrix based on semantic distance measures as described in Section 3.3 and the predefined number of clusters represent the input to the spectral clustering algorithm presented as Algorithm 1. Based on the input matrix a degree matrix is formed. The difference between the degree and the weighted adjacency matrix forms the graph Laplacian L. The normalized matrix of eigenvectors of the normalized L is then used as input to the k-means algorithm. We followed Von Luxburg (2007) and tested the ϵ-neighborhood, k-nearest neighbor, and a fully connected graph on our dataset. The algorithm provides the number of clusters that was initially provided as input. To optimize this variable, we experimented with different sizes of k detailed in the Section 4.

Algorithm 1 Normalized Spectral Clustering (Ng et al., 2001)

```
Input: Similarity matrix $S \in \mathbb{R}^{nxn}$, number of $k$ clusters
Construct a degree matrix $D$ where $d_{ij}$ = $\sum_{j=1}^{n} w_{ij}$ and $d_{ij}$ = 0 if i $\neq$ j
Construct a similarity graph and its weighted adjacency matrix $W$
Construct a graph Laplacian $L = D - W$
Compute the normalized Laplacian $L_{sym} := D^{-1/2}LD^{-1/2}$
Compute the first $k$ eigenvectors $v_1$,...$v_K$ of $L_{sym}$ and write them as columns into the matrix $U$ $\in \mathbb{R}^{nxk}$
Compute the matrix $T \in \mathbb{R}^{nxk}$ from $U$ by normalizing that is set $t_{ij}$ = $u_{ij}/(\sum_k u_{ik}^2)^{1/2}$
Let $y_i$ be the vector corresponding to the $i^{th}$ row of $T$
Cluster the points $(y_i)_{i=1,...,n}$ with the k-means algorithm into clusters $C_1$,...,$C_k$
Output: Clusters $C_1$,...,$C_k$ with $C_i$ = $\{j|y_j \in C_i\}$
```

3.5 Semantic Role Labeling

Since we obtained a large number of clusters from experimenting with different cluster sizes k, different similarity matrices, and different methods to transform the similarity matrices to graphs, we were looking for an automated method to evaluate the resulting clusters in terms of their spatial schemas. To this end, we employ a semantic role labeling tool called Curator (Punyakanok et al., 2008), which follows the notation of the PropBank project Kingsbury & Palmer (2002). We input each verb-preposition-noun triple to Curator and extract the label assigned to the preposition. Each verb-preposition combination has a different number of nouns assigned to it depending on their cooccurrences in the corpus. We assign each verb-preposition combination with the most frequently occurring label in the results from the semantic role labeling. In case frequencies of two labels are identical, we favor spatial labels, namely: *Location*, *Journey*, *PhysicalSupport*, *EndState*, *Destination*, *StartState*, and *Source*, over non-spatial labels. In a second step, we represent each cluster as a collection of those role labels with their corresponding frequencies. This process allows us to determine the most interesting combination of cluster size, similarity matrix, graph building, and clustering algorithm in the sense of representing the most "purely spatial" clusters. This method is based on the assumption that clusters that are predominantly spatial are good indicators for image-schematic structures.

3.6 Image Schema Identification

After assigning semantic role labels to each cluster, we analyze the spatial subset of the cluster collection regarding potential image-schematic content. To identify image schemas, two experts manually analyzed the content of each cluster using definitions of image schemas and their features as reference material introduced by Johnson (1987), Lakoff (1987), and Kövecses (2010). Here image schemas are described through their salient features. For instance, CONTAINMENT is described as having an *inside*, an *outside*, and a *boundary* and VERTICALITY requires *movement* or *directionality* as either *up* or *down*. To exemplify this approach, let us look at the triple "bring-into-disrepute". It describes a transformation from the state of good, or neutral, reputation to a different more negative reputation, the *disrepute*. There is a clear boundary between those two states and certain events may cause this state to change, in this case an event *brings* about this transformation. While this is an abstract transformation of one state to another, conceptually it runs in parallel to a concrete situation in which an object moves into a container. Thus, this triple was identified as CONTAINMENT in the manual analysis of clusters for image schemas. More examples are provided in Section 4. During the detection of image schemas, all experts performing the task were given access to the same image schema definitions to refer to when in doubt on the image schematic nature of the clusters.

4. Results

Due to the size of the corpus, we obtain a large collection of potential image-schematic clusters. In Section 3, we explain how spatially relevant and image-schematically sound clusters are detected from this collection. This section presents the relevant statistical measures that reveal which combination of settings worked best and which image-schematic structures we detected.

4.1 Clustering and Role Labeling Results

Spectral clustering takes a similarity matrix and the number of clusters as input. To generate the similarity matrix we experiment with three different semantic distance measures: Jensen-Shannon Divergence (JSD), Positive Pointwise Mutual Information (PPMI), and Term-Frequency Inverse Document Frequency (TF-IDF) presented in Section 3.3. In terms of preset numbers of clusters, we experiment with sizes 50, 100, 200, and 300. To transform the similarity matrix into a graph, we use three different methods following the guidance in Von Luxburg (2007): ϵ-neighborhood graph (ϵ), k-nearest neighbours (knn), and fully connected graphs (fc). For instance, for similarity matrix PPMI, cluster size $k = 300$ and knn as a graph building method, we obtain a verb-preposition cluster of average size 11.8, such as the following:

Example 1: Cluster $k = 300$, PPMI, knn

```
Cluster 55: ['bring-into', 'bringing-in', 'bringing-into',
'brings-into', 'brought-in', 'brought-into', 'fall-in', 'falls-in']
```

No linguistic preprocessing in terms of lemmatization or stemming is required, since inflectional variants such as *bringing*, *brings*, *brought*, and *bring* are clustered together as shown in Example 1. We also decided against lemmatization since we were interested in whether different grammatical tenses were grouped together. This is interesting because image-schematic structures are usually described as spatio-temporal schemas, so when coming from natural language data, grammatical tenses are important.

From the Europarl corpus, we derived 97,211 unique verb-preposition-noun triples (1,636,291 with all instances regardless of duplicates). On occasion the dependency parsing provided false positives, such as interpreting 'p.m.' as a noun. Those false positives are inevitably part of our cluster collection. However, one of the advantages of this approach is that they are usually grouped together and thus easy to exclude. During the first clustering process, we detected that linkers (e.g. "for example" or "on the one hand") resulted in misleading triples that did not contain any actual image schemas or spatial language. This is because linking devices are discourse structuring elements with pre-specified non-spatial semantics. Thus, we decided to exclude all linkers when providing the input to the spectral clustering algorithm.

On top of the three similarity matrices (PPMI, JSD, TF-IDF), the four cluster sizes (50, 100, 200, 300), and the three graph building methods (knn, fc, ϵ), we also experimented with three different algorithms for clustering: normalized spectral clustering according to Ng et al. (2001), normalized spectral clustering according to Shi & Malik (2000), and unnormalized spectral clustering. Here we present only the results from the normalized algorithm by Ng et al. (2001), as it provided substantially better results than the other two. We also exclude the results from the size 50 clusters and TF-IDF similarity matrix in the following presentation as they provided the least interesting results.

To compare the different settings, we perform semantic role labeling on all triples. For instance, the verb-preposition pair *bring-into* in Cluster 55 co-occurred with the nouns represented in Example 2 (the numbers represent the frequency of occurrence of each triple). The role labels we obtained for this verb-preposition combination were *Destination=8, EndState=1*, where 'bring-into-contact' was labeled as 'EndState'.

Example 2: Feature vector of 'bring-into' of Cluster 55

```
bring-into: {'disrepute': 15, 'play': 29, 'force': 36, 'focus': 18,
'contact': 11, 'question': 21, 'operation': 14, 'effect': 18,
'european union': 15}
```

Table 1. Comparison of Setting Combinations for Clustering

Algorithm method: Cluster size:	PPMI knn	PPMI fc	PPMI ϵ	JSD knn	JSD fc	JSD ϵ
100	29%	22%	24%	18%	20%	18%
200	21%	22%	22%	6%	23%	19%
300	**31%**	29%	24%	8%	23%	24%

This consequently means in our method that the overall label for this verb-preposition combination is *Destination*. In the end, we obtain a representation of all verb-preposition combinations in a cluster by means of labels, e.g. *Destination=7, EndState=1* for the 8 combinations in Cluster 55 in Example 1. This example also nicely shows the combination of concrete, e.g. 'European Union', and abstract, e.g. 'disrepute', concepts in each cluster. We then analyze each cluster for purity of spatial tags as described in Section 3.5. In Table 1 we present the results for the normalized algorithm by Ng et al. (2001) for cluster sizes 100, 200, and 300, similarity metrics PPMI and JSD, and all three graph building methods knn, fc, and ϵ.

As can be seen in Table 1 the combination of cluster size $k = 300$, PPMI, and knn has the highest percentage of 'pure' spatial clusters with 31%. Mixed clusters contain spatial and non-spatial labels and are 16% of all clusters, while the remaining 53% of the clusters mainly contain labels other than those specified as spatial in Section 3.5. The image schema detection results below are based on those 300 clusters from this combination with on average 11.8 verb-preposition combinations in each cluster.

To measure the accuracy of the semantic role labeling, we manually evaluated the 300 clusters to their spatial content. We obtain 126 spatial clusters out of 141 total clusters with spatial labels, that is, considering 'pure' and 'mixed' clusters. Out of all 300 clusters a total of 145 are spatial, which results in a precision of 89.36%, a recall of 86.90%, and an F-Measure of 88.11% for our semantic role labeling evaluation.

4.2 Image Schema Identification Results

From the 300 clusters, a total of 92 (31%) were tagged as purely spatial, 49 (16%) as mixed spatial and other labels, and 159 (53%) contained no spatial labels. The detection of image-schematic structures was conducted manually by two experts with an inter-rater agreement of 78% on the 92 spatial clusters. For the 20 clusters that were not assigned the same image schema by the two experts, a third expert was consulted. Table 2 presents the image schemas as detected by the majority of the experts. The overall precision achieved with this method is 80.43% of the 110 schemas found across all 300 clusters and the recall is 67.27%, with an F-Measure of 73.27%.

The predominant image-schematic structure we detected in the spatial clusters after following the technique described in Section 3.6, was CONTAINMENT. Table 2 presents quantified results, where 'Freq. 40' refers to the number of clusters identified as CONTAINMENT. 'Other' in the image schema column refers to the collected occurrences of the image schema structures NEAR-FAR, SPLITTING, PART-WHOLE, and CENTER-PERIPHERY, each of which occur only once. The two

Table 2. Detected Image Schemas

	Pure Clusters		Mixed Clusters		Other Clusters	
Image Schema	Freq.	Percent	Freq.	Percent	Freq.	Percent
CONTAINMENT	40	43%	12	24%	9	6%
SOURCE_PATH_GOAL	18	20%	2	4%	2	1%
SUPPORT	8	9%	3	6%	5	3%
SURFACE	2	2%	-	-	2	1%
VERTICALITY	2	2%	1	2%	-	-
Other	4	4%	-	-	-	-
Total Schemas	**74**	**80%**	18	37%	18	11%
Total Clusters	92	100%	49	100%	159	100%

columns under ‘pure clusters’ show the number and types of image schemas for clusters labeled solely with spatial tags. These are the most interesting ones since they show how effective our method is. Both in the ‘mixed clusters’, where the role labels are not purely spatial, as well as in the ‘other clusters’ that are exclusively assigned non-spatial labels by the role labeling process, we found 18 image schemas respectively.

While CONTAINMENT was predominantly detected by the prepositions *in* and *within*, some clusters demonstrated interesting combinations in which several different CONTAINMENT structures were included. For example, cluster 63 had both ’excluded-from’ and ’integrate-into’ as present verb-preposition combinations. One further example of CONTAINMENT is the triple “fall-in-love” obtained for example from the sentence fragment “...the South Caucasus ... is a region with which one quickly falls in love”, which is also part of Cluster 55 in Example 1. Another example is “play-within-framework” taken for instance from the fragment “the role the European Parliament will have to play within this framework”. The second most frequent cluster group relating to the schema SOURCE_PATH_GOAL contained a wide variety of prepositions, such as *by*, *at*, and *along*. With this schema the combination of abstract and concrete concepts can also be nicely exemplified. For instance, the sentence “Turkey must be encouraged to *continue along this road*” is a rather literal and concrete use of SOURCE_PATH_GOAL, whereas others are more abstract, such as “I definitely believe that Europe must *start from scratch*”.

5. Discussion

Previous approaches to image schema extraction are mostly theory-driven and use hand-crafted examples (e.g. Dodge & Lakoff (2005)) or lexico-syntactic patterns (e.g. Bennett & Cialone (2014)). While this method has facilitated the uncovering of new image-schematic structures in language (see e.g. Gromann & Hedblom (2016)), their reliance on linguistic surface structures and poor portability to other languages and domains makes them undesirable methods to use. To overcome these limitations, we present an effective clustering-based method to detect image-schematic structures. This method can be ported to all languages for which there exists a sufficiently accurate dependency parser that allows for the extraction of prepositions and their related nouns and verbs. The results from the evaluation show that the proposed method identifies spatial relations regardless of the surface structure of expressions. In other words, the method finds spatial relations independent of particular prepositions or words. For example, Cluster 63 demonstrates several spatio-temporal relations that are included in the CONTAINMENT schema, that is, 'excluded-from' and 'integrate-into' demonstrate being on the outside.

One could argue that the semantic role labeling is already sufficient to detect spatial language. However, several prepositions in our dataset were not assigned a tag during the labeling process, such as "within" and "below", which we labeled manually. In contrast, the approach presented in this paper does not rely on previous annotations and successfully grouped triples with "within" together with other combinations of the same meaning, as was the case for "below". In addition, role labeling alone would not be able to group different verb-preposition combinations with non-identical spatial labels based on their similar spatial semantics, such as shown above for Cluster 63 and Cluster 55.

In terms of limitations, the proposed method currently fully relies on nouns as features for the clustering algorithm. It would be interesting to see the difference in results if a richer feature set was used for the clustering process. For instance, it would be interesting to experiment with features for clustering using the semantic roles of verbs and nouns, either in solitude or in a combination with the semantic roles of the preposition, or using additional distributional features. One of the biggest limitations of our approach is the necessity to manually identify which image schema a cluster represents. This could potentially be overcome by using a supervised approach, clustering or classification, based on examples of already identified image schemas in natural language as provided by this approach for English. Another alternative is to submit the clusters ready for image schema detection to a crowdsouring platform. A major challenge with crowdsourcing image schemas and their detection is the difficulty to explain their nature to non-experts in few words.

The method proposed in this paper for the time being focuses on the extraction of image schemas from natural language text without considering the highly common combination of image schemas within a single expression. For instance, the expression *'to get into trouble'* represents a CONTAINER as much as a PATH. Since this to the best of our knowledge is the first proposal of a semi-automated method for image schema detection, we decided to focus as a first step on the identification of one image schema per cluster. While this could theoretically lead to a low agreement of annotators, we found the opposite to be true in that our annotators agreed in 78% of all clusters on the kind of image schema and not just on the fact that the cluster represented an image schema.

Image schemas provide vital information of concepts and events expressed in language and by identifying them it is made possible to (eventually) integrate them into artificial systems. Here we believe it is possible to bypass some of the complicated formal representations currently needed (as demonstrated above with the 'egg cracking' example (Morgenstern, 2001)). This could be used in human-agent interfaces focusing on concrete spatial navigation, or even in systems that tackle a more abstract understanding of natural language such as an analogy engine or a system for computational concept invention (Hedblom et al., 2016). Our method has shown that basic sensorimotor concepts needed for manipulation instructions can be obtained directly from language, such as abstract and concrete CONTAINERS as well as movement along a PATH. This represents a first step towards detecting kinesthetic early experiences in abstract mental concepts as manifested in natural language. However, one problem that still remains is the formal representation of image schemas. While there exists some formalization approaches to image schemas (e.g. (Kuhn, 2007; Hedblom et al., 2015; Bennett & Cialone, 2014)) there is much work still remaining before this is a feasible project.

6. Conclusion and Future Work

Natural language understanding in cognitive systems requires a detailed analysis of the semantics underlying language. To contribute to the analysis of how language, real-world objects, and mental representations interact, this paper introduces a method to semi-automatically identify image schemas in natural language using spectral clustering and semantic role labeling. The results are promising and show that the method can be used effectively to identify image-schematic structures. We intend to further improve it by experimenting with different feature sets.

Our method extracts a large number of verb-preposition-noun triples that are annotated with image schemas. This represents a powerful repository that can be used in linguistic analyses of spatial language and conceptual metaphor detection as well as prove useful for other machine learning techniques aiming for image schema extraction as it provides examples for supervised approaches. We also see potential uses of such a repository by robotics and AI systems requiring natural language understanding. However, for that purpose more fine-grained formal representations of the detected image schemas are required.

Image schemas have been claimed to be universal, a claim that our future work intends to challenge by extending our research to other languages and domains. This at the same time would prove the portability of our method. As a final note, we are optimistic that this cognitively inspired method has potential to improve natural language understanding, if successfully integrated into a formal framework.

References

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. *ACL (1)* (pp. 238–247).

Bennett, B., & Cialone, C. (2014). Corpus guided sense cluster analysis: a methodology for ontology development (with examples from the spatial domain). *8th Int. Conf. on Formal Ontology in Information Systems (FOIS)* (pp. 213–226). IOS Press.

Chen, D., & Manning, C. D. (2014). A fast and accurate dependency parser using neural networks. *EMNLP* (pp. 740–750).

Dodge, E., & Lakoff, G. (2005). Image schemas: From linguistic analysis to neural grounding. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: Image schemas in cognitive linguistics*, 57–91. Berlin: Mouton de Gruyter.

Feldman, J., Dodge, E., & Bryant, J. (2009). A neural theory of language and embodied construction grammar. In B. Heine & N. Narrog (Eds.), *Oxford handbook of linguistic analysis*, 111–138. Oxford University Press.

Gibbs, R. W., & Colston, H. L. (1995). The cognitive psychological reality of image schemas and their transformation. *Cognitive Linguistics*, *6*, 347–378.

Gromann, D., & Hedblom, M. M. (2016). Breaking down finance: A method for concept simplification by identifying movement structures from the image schema path-following. *Proc. of the Joint Ontology Workshops (JOWO)*.

Guerin, F. (2008). Learning like a baby: A survey of AI approaches. *The Knowledge Engineering Review*, *00*, 1–22.

Hampe, B., & Grady, J. E. (2005). *From perception to meaning: Image schemas in cognitive linguistics*, volume 29 of *Cognitive Linguistics Research*. Berlin: Walter de Gruyter.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*, 146–162.

Hedblom, M. M., Kutz, O., & Neuhaus, F. (2015). Choosing the right path: image schema theory as a foundation for concept invention. *Journal of Artificial General Intelligence*, *6*, 22–54.

Hedblom, M. M., Kutz, O., & Neuhaus, F. (2016). Image schemas in computational conceptual blending. *Cognitive Systems Research*, *39*, 42–57.

Johnson, M. (1987). *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reasoning*. The University of Chicago Press.

Kingsbury, P., & Palmer, M. (2002). From treebank to propbank. *LREC* (pp. 1989–1993). Citeseer.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. *MT summit* (pp. 79–86).

Kollar, T., Tellex, S., Roy, D., & Roy, N. (2014). Grounding verbs of motion in natural language commands to robots. *Experimental robotics* (pp. 31–47). Springer.

Kordjamshidi, P., Van Otterlo, M., & Moens, M.-F. (2011). Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Transactions on Speech and Language Processing (TSLP)*, *8*, 4:1–36.

Kövecses, Z. (2010). *Metaphor: A practical introduction*. Oxford University Press, USA.

Krishnamurthy, J., & Kollar, T. (2013). Jointly learning to parse and perceive: Connecting natural language to the physical world. *Transactions of the Association for Computational Linguistics*, *1*, 193–206.

Kuhn, W. (2007). An Image-Schematic Account of Spatial Categories. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers (Eds.), *Spatial information theory*, volume 4736 of *LNCS*, 152–168. Springer.

Lakoff, G. (1987). *Women, fire, and dangerous things. what categories reveal about the mind*. The University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. The University of Chicago press.

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. Basic Books.

Litkowski, K. (2014). Pattern Dictionary of English Prepositions. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (pp. 1274–1283). Baltimore, Maryland.

Litkowski, K. C., & Hargraves, O. (2005). The preposition project. *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications* (pp. 171–179).

Mandler, J. M., & Pagán Cánovas, C. (2014). On defining image schemas. *Language and Cognition*, (pp. 1–23).

Misra, D. K., Sung, J., Lee, K., & Saxena, A. (2016). Tell me dave: Context-sensitive grounding of natural language to manipulation instructions. *The International Journal of Robotics Research*, *35*, 281–300.

Morgenstern, L. (2001). Mid-Sized Axiomatizations of Commonsense Problems: A Case Study in Egg Cracking. *Studia Logica*, *67*, 333–384.

Ng, A. Y., Jordan, M. I., Weiss, Y., et al. (2001). On spectral clustering: Analysis and an algorithm. *NIPS* (pp. 849–856).

Papafragou, A., Massey, C., & Gleitman, L. (2006). When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition*, *98*, B75–B87.

Pauwels, P. (1995). Levels of metaphorization: The case of put. In L. Goossens (Ed.), *By Word of Mouth: Metaphor, metonymy and linguistic action in a cognitive perspective*, 125–158. Amsterdam: John Benjamins Publishing Company.

Punyakanok, V., Roth, D., & Yih, W. (2008). The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, *34*.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, *22*, 888–905.

Shutova, E., Sun, L., Gutierrez, D., Lichtenstein, P., & Narayanan, S. (2016). Multilingual metaphor processing: Experiments with semi-supervised and unsupervised learning. *Computational Linguistics*. Forthcoming.

Sun, L., & Korhonen, A. (2009). Improving verb clustering with automatically acquired selectional preferences. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2* (pp. 638–647). Association for Computational Linguistics.

Talmy, L. (2005). The fundamental system of spatial schemas in language. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: Image schemas in cognitive linguistics*, volume 29 of *Cognitive Linguistics Research*, 199–234. Walter de Gruyter.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, *17*, 395–416.

Xu, Z., & Ke, Y. (2016). Effective and efficient spectral clustering on text and link data. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 357–366). ACM.

Zlatev, J. (2010). Spatial semantics. In D. Geeraerts & H. Cuyckens (Eds.), *The oxford handbook of cognitive linguistics*, 318–350. Oxford University Press.