# Body-Mind-Language: Multilingual Knowledge Extraction Based on Embodied Cognition

Dagmar Gromann[1] and Maria M. Hedblom[2]

[1] Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Spain
[2] Free University of Bozen-Bolzano, Bozen-Bolzano, Italy

**Abstract** Cognitive linguistics has provided compelling evidence that semantic structure in natural language reflects conceptual structure that arises from our embodied experience in the world. To capture this conceptual structure, a set of spatio-temporal cognitive building blocks called image schemas was introduced by Lakoff and Johnson. Detecting image schemas in natural language can provide further insights into how embodied experiences are encoded in natural language and potentially contribute to research on conceptual understanding and symbol grounding in cognitive systems. Methods for (semi-)automatically extracting image schemas from natural language are an open challenge. We propose a spectral clustering approach paired with semantic role labeling to semi-automatically extract image schemas from multilingual text, obtaining a precision of more than 80% on three languages.

## 1 Introduction

Embodied agents and cognitive systems need to learn spatio-temporal knowledge in order to be able to interact with the world. Cognitive linguistics and psychology have introduced convincing evidence that spatio-temporal relations are highly frequent in natural language [17,37,39]. This is why language represents an important source for embodied agents to learn spatial relations [1], e.g. for human-robot interaction [16,27], making agents learn like a baby [10], and basically any mapping between symbols and objects in the physical world [19]. However, the symbol grounding problem of how signs are assigned with meaning, relate to real-world objects, and are cognitively represented remains open.

While human cognition still is a mystery on many levels the latest paradigm shift in the view of cognition, embodied cognition, provides some new promising insights. Embodied cognition states that all cognition occurs as a consequence of the body's sensorimotor experiences with the environment [34]. Image schemas are introduced within the context of embodied cognition and described as spatio-temporal relations that infants learn in the early years through repeated exposure to particular events. For example, the image schema CONTAINMENT is learnt when the child is repeatedly exposed to objects moving in and out of containers [24].

Extracting image schemas from natural language can provide a principled way to investigate the connection of thought and language and gain new insights into the cognitive grounding of natural language. To the best of our knowledge, automatically detecting image schemas in natural language is an open challenge. Until recently, many linguistically-relevant studies on image schemas have focused on the lexical surface structure of expressions [3,8] or provided manually curated examples [6,12] to support their claims.

In this paper, we address this challenge by proposing a semi-automated method to detect image schemas in three languages based on unsupervised spectral clustering paired with semantic role labeling, extending our previous work on methods in English [9]. The method builds on findings from research on spatial language (e.g [17,39,42]) in which prepositions were utilized as spatial indicators and verbs as indicators for motion and temporal change. In the proposed method verb-preposition pairs are clustered with co-occurring nouns as features and their relative frequencies as feature values. For instance, for "continue-along" the feature vector is [*'route': 13, 'road': 37, 'path': 94, 'lines': 53*]. Verb-prepositions pairs are clustered based on their feature vectors. The above example is grouped in the cluster [*'continue-along', 'continuing-along', 'progress-along', 'set out-on', 'set-on'*]. We use existing automated tools for semantic role labeling to separate clusters into spatial or non-spatial based on the main prepositions senses, e.g. DIRECTION for the majority of prepositions in the above cluster which we consider spatial. All clusters are then manually annotated with image schemas, e.g. SOURCE_PATH_GOAL in the above example. The outcome is a repository of verb-preposition clusters with their feature nouns and their original sentences in English, German, and Swedish annotated with role labels and identified image schemas that carries the potential of improving spatial language understanding and its cognitive grounding.

The remainder of this paper is structured as follows. First, we introduce image schemas and some related approaches on extracting spatial information from text. Second, we describe the utilized dataset and the proposed method. Third, the results of the method are presented as well as a summary of the identified image schemas across the three languages. Finally, we discuss the results and in the conclusion we include a few remarks on future work.

## 2 Foundation

As cognitive research started to focus on the sensorimotor experiences as the foundation of cognition, rather than the classical *cognitivist* view (i.e. 'cognition is computation') that previously dominated cognitive science [34], image schemas were introduced by Lakoff [20] and Johnson [13] as a theory to explain certain cognitive phenomena, in particular the conceptualisation of concepts in terms of spatial language. While image schemas evolve from concrete sensorimotor experiences, their mental representation is considered abstract. Psychological support for image schemas comes from how they offer infants conceptual grounds to make predictions about their surroundings [24,7]. Indeed, work in linguistics (e.g. [6]) and psychology (e.g. [24]) 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 [21]. 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 of how concrete sensorimotor experiences are transferred to abstract adult communication. Pauwels [30] went so far as to claim that any abstract use of the word "put" requires the understanding of CONTAINMENT stressing the importance of verbs in image schema analyses.

One semi-automated method designed to extract spatial relations between a trajector and a landmark was introduced by Kordjamshidi et al. [17]. Their method uses machine learning on word triples by connecting 'the trajector' and 'the landmark' through a preposition, or as they call it 'a spatial indicator'. Prepositions have been demonstrated to be essential in terms of revealing spatial information [22], yet they do not always capture motion and temporal change. For this also verbs need to be taken into account as they play a central role in identifying the relation between trajectors and landmarks [16], which, however, is not the case in Kordjamshidi et al. [17]. Approaches that have taken verbs into account, such as [25], instead often rely on handcrafted rules or predefined spatial expressions to extract motion verbs across languages. Most work on extracting image schemas from natural language has either been done by curating manual examples [6] or by conducting corpus studies based on specific lexico-syntactic patterns [3,8]. As image schemas tap into the core of conceptual metaphors [21,26], their identification in language also opens up new possibilities in not only identifying the concrete, but also the abstract. But for this it is necessary to abstract away from the lexical surface structure of expressions, which is not possible with lexico-syntactic patterns. For instance, the expression "to have an empty life" characterizes life as having a feature that can either be full or empty, directly transferred from physical characteristics of CONTAINERs, which, however, could never be found when querying for specific verbs or nouns. We methodologically benefit from this connection to metaphors by building on clustering approaches for multilingual metaphor detection [36].

## 3 Dataset

To comprehensively evaluate the occurrences of image schemas in natural language as well as to efficiently perform clustering, a large natural language corpus is needed. Additionally, in order to make the results comparable across different languages, the corpus needs to be parallel. Thus, we decided to use the Europarl corpus [15], which is aligned across several languages, commonly used in linguistic approaches, and with its 1,959,830 sentences in English, large enough for our purposes. The corpus contains sentences extracted from the proceedings of the European Parliament, meaning that the coverage of the corpus is somewhat limited to topics revolving around governance and political issues.

## 4 Methodology

One established exploratory data analysis method is unsupervised clustering [2]. The chosen normalized spectral clustering algorithm proposed by Ng et al. [28] has been effectively applied to various lexical acquisition tasks (e.g. [36,41,38] ). We build on successful methods for conceptual metaphor extraction [36] in combination with findings from spatial language analysis (e.g. [39]) for our method. The combination of verbs and prepositions as indicators for spatial and potentially image-schematic structures, is backed by manual corpus-based analyses on image schema detection (e.g. [6]). This section describes the individual steps from text to image schema depicted in Fig. 1.

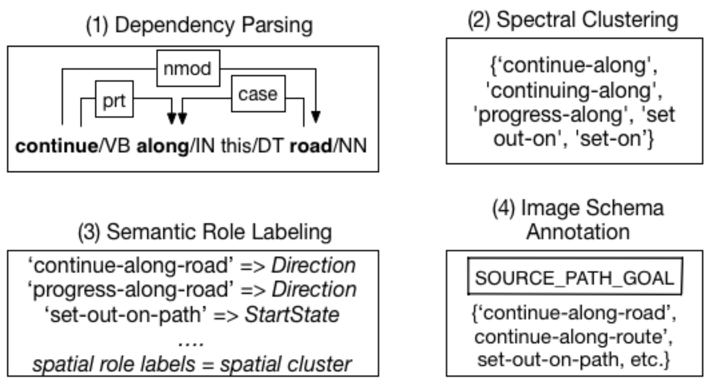


**Figure 1.** Individual steps of proposed method

### 4.1 Dependency Parsing

Dependency parsing identifies how individual sentential elements depend on each other by first part-of-speech (POS) tagging each element and then analyzing the structure of the whole sequence. We parse the dataset for each language and extract verb-preposition-noun combinations. From each sentence, such as (1), we first extract the preposition ("along"), its dependent noun ("road") or noun phrases and search through all their dependency relations for a verb ("continue") and potential phrasal particle as depicted in Step (1) of Fig. 1.

(1) This is why Turkey must be encouraged to continue along this road.

Thereby, we obtain verb-preposition pairs and all the co-occurring nouns with their relative frequencies, which represent the feature vector for our clustering algorithm. Only verb-preposition pairs that occurred at least ten times with the extracted noun where considered for the clustering to avoid a distortion of the clusters by rare words or dependency parsing errors.

For English and German we used the Stanford Dependency Parser [5]. However, Swedish is not available in the Stanford tool set so we chose Stagger [29] for POS tagging and the data-driven parser-generator MaltParser [11] for dependency parsing. We use the Swedish MaltParser Model (swemalt-1.7.2.mco) that was trained on the Talbanken section of the Swedish Treebank.

### 4.2 Spectral Clustering

Spectral clustering is particularly attractive since it is reasonably fast and transforms the data clustering into a graph partitioning problem, partitioning based on the values of the edges. It takes a similarity matrix and the number of clusters as input. Computing a similarity matrix depends on the choice of semantic distance measure that is best for the given data. We tested on Term Frequency-Inverse Document Frequency (TF-IDF) as one of the most common similarity measures, Positive Pointwise Mutual Information (PPMI), and the Jensen-Shannon divergence (JSD), a symmetric and a smoothed version of the Kullback-Leibler that has successfully been used in conceptual metaphor clustering [36]. We evaluated each of those similarity measures by generating similarity matrices and submitting them to the algorithm. The results of this process were analyzed by semantic role labeling and success was defined as the best separation of information into spatial and non-spatial clusters.

We finally used PPMI for the image schema annotation (detailed in Section 5), which is why we only discuss this metric further. 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 $x$ representing the frequency of a verb-preposition pair and $y$ representing a noun it co-occurs with, 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.

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

$$PPMI = \begin{cases} PMI(x, y), & \text{if } PMI(x, y) > 0 \\ 0, & \text{otherwise} \end{cases}$$

The similarity matrix captures the semantic distance between verb-preposition pairs based on their co-occurring nouns, which represent the features of the algorithm. We tested on unnormalized, normalized according to Shi and Malik [35] and normalized according to Ng et al. [28] spectral clustering. Algorithm 1 presents the most successful normalized algorithm by Ng et al. [28], where success is defined as balanced cluster sizes and correct separation of spatial and non-spatial information. The major problem with the Shi and Malik normalization [35] for our approach was the strong variation in cluster size, leading at times to clusters of more than 2,000 verb-preposition pairs.

In Algorithm 1 the difference between the degree and the weighted adjacency matrix leads to 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 [40] and tested the $\epsilon$-neighborhood, k-nearest neighbors, and a fully connected graph building methods on our dataset. The algorithm outputs the number of clusters that was initially indicated. To optimize this variable, we experimented with different sizes of $k$ detailed in Section 5. Our assumption for this method was that verb-preposition pairs co-occurring with similar nouns/noun phrases might exhibit a similar spatio-temporal behavior. Each cluster groups verb-preposition pairs based on that similarity.

**Algorithm 1** Normalized Spectral Clustering [28]

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

### 4.3 Semantic Role Labeling

From experimenting with similarity matrices (TFIDF, PPMI, JSD), cluster sizes $k$ (50, 100, 200, 300), and clustering algorithms we obtained 3,900 clusters for each language. This called for an automated method to compare the resulting clusters. We found that semantic role labeling can effectively be used to separate spatial from other verb-preposition pairs for each cluster, thus providing us with a first purity estimation of the individual clusters. Semantic role labeling abstracts away from syntactic variation and assigns labels to arguments of sentence predicates by means of predefined relations.

In semantic role labeling, the determiner for spatial information is the preposition sense that is assigned based on the verb and noun the preposition relates to. Thus, we take the verb-preposition pairs of each cluster and query existing tools for the preposition sense using the feature nouns of each pair, e.g. “continue-along-road” is a DIRECTION. Second, we accumulate all labels obtained for a specific verb-preposition pair, e.g. one of them being the above DIRECTION with the noun “road” for “continue-along”. The most frequent label of those accumulated labels is assigned to the pair. We classified all preposition senses as either spatial or non-spatial. If most verb-preposition pairs in a cluster have spatial labels, we consider the whole cluster spatial such as the one in Fig. 1. As the example shows, we did not perform lemmatization (“continue” and “continuing” are in the cluster) and we considered phrasal verbs, such as “set out” as well as noun phrases.

The evaluation of the purity of a cluster is estimated based on verb-preposition role labels and their frequency in a cluster. We differentiate between spatial (>80% spatial labels), mixed (>30% spatial labels), and other (<30% spatial labels) clusters. Semantic role labeling is language-specific, which means we had to find different solutions for each language.

For English, we employed a semantic role labeling tool called Curator [31], which follows the notation of the PropBank project [14] and its fine-grained labeling of preposition roles. Curator provided highly accurate as well as detailed semantic role labels for the whole English dataset, including annotations for verb, noun and preposition senses. The preposition sense annotation was a main motivator for choosing Curator, since other tools frequently only tag prepositions as "prepositions" without specifying their sense in detail. We classify all role labels into either spatial or non-spatial, where the former for our case are: *Location*, *StartState*, *EndState*, *Source*, *Destination*, *Direction*, *PhysicalSupport*, and *Journey*.

For Swedish we relied on the preposition senses provided with the Swedish Treebank model of MaltParser, which differentiates between spatial (RA), temporal (TA), and several other types of adverbials, providing a less fine-grained and less accurate, but still viable, estimation of which cluster setting to analyze for image schemas.

For German, semantic role labeling is a challenging endeavor since many parsers [4,32] focus on valency-bound complements. We could not find any parser that provided equivalent preposition sense labeling results as for English and Swedish and thus decided to annotate the list of verb-preposition-noun triples extracted from text manually differentiating spatial and non-spatial labels. This manual annotation was then used to evaluate the cluster settings.

### 4.4 Image Schema Annotation

In a final step we analyzed the resulting clusters for their image-schematic content. Our definition of image schemas was based on definitions obtained from Johnson [13], Lakoff [20], and Kövecses [18] as exemplified in Table 1.

**Table 1.** Example of image schemas and definitions

| Image Schema | Description |
|---|---|
| CONTAINMENT | Boundary, enclosed area or volume, or excluded area or volume [13] |
| SOURCE_PATH_GOAL | Source or starting point, goal or endpoint, a series of contiguous locations connecting those two, and movement [13,20] |
| SUPPORT | Contact between two objects in the vertical domain [23] |

Each cluster was manually analyzed based on those definitions exemplified in Table 1 to determine which image schema, if any, was the most dominant in the cluster. For English, two annotators separately assigned image schemas and upon disagreeing, a third annotator took the final decision. For Swedish and German there was only one annotator available for each language, resulting in less reliable results than for English.

Consider, for instance, the triple "bring-into-disrepute" that describes a transformation from the state of good, or no, reputation to a negative reputation, *'disrepute'*. Despite being abstract, there is a clear boundary between the two states and certain events may cause this state to change, in this case an event *'brings'* about this transformation. It can be argued to correspond to the movement *'into'* a CONTAINER. Annotators evaluated each verb-preposition-noun triple in a cluster to decide whether it represents any image-schematic structure.

## 5 Results

Due to the size of the corpus, we obtain a large collection of potentially image-schematic clusters. In Section 4, we explained how spatially relevant and image-schematic clusters are detected from this collection. This section presents quantified results of the best combination of settings and the number of obtained image-schematic structures. Dependency parsing is taken as given even though some verb-preposition pairs might have been overlooked by the parser and we start with describing the clustering results.

### 5.1 Clustering Results

We obtained 3,900 clusters from three similarity metrics (JSD, PPMI, TFIDF) to build the similarity matrices, three algorithms (unnormalized, two normalized), three graph building methods (knn, $\epsilon$, fully-connected), and four cluster input sizes $k$ (50, 100, 200, 300) for each language. A comparison in English is illustrated in Table 2 with normalized clustering [28] to exemplify the selection process of the best settings since space does not permit a detailed description for all languages. A total of 92 (31%) were tagged as purely spatial, 49 (16%) as mixed spatial and other labels, and 159 (53%) contained less than 30% spatial labels with knn, PPMI, and k of size 300. This represented the highest number of spatial clusters, which is why we analyzed this combination of settings for image schemas.

While knn clearly returned the best results for all languages, in German JSD and TF-IDF with normalized clustering by Shi and Malik [35] returned higher numbers than PPMI, especially for size 50 clusters. However, upon manually inspecting the clusters, the clusters turned out to be very large and contain highly mixed information. Thus, we manually inspected ten clusters of each setting combination, which returned the best results for the same settings as for English. In Swedish, the role frequency-based approach equally pointed to other combinations with inferior quality, and also for Swedish the best settings turned out to be the same as for English.

**Table 2.** Comparison of setting combinations for English clustering

| Algorithm method:<br>Cluster size (100-300) | PPMI<br>knn | PPMI<br>fc | PPMI<br>$\epsilon$ | JSD<br>knn | JSD<br>fc | JSD<br>$\epsilon$ |
|---|---|---|---|---|---|---|
| 100 | 29% | 22% | 24% | 18% | 20% | 18% |
| 200 | 21% | 22% | 22% | 6% | 23% | 19% |
| 300 | **31%** | 29% | 24% | 8% | 23% | 24% |

We clustered a total of 2,259 English, 3,234 Swedish, 2,739 German unique verb-preposition pairs based on their feature vectors with an overall frequency above ten occurrences in the corpus. The average cluster size for English was 10.40 verb-preposition pairs, for Swedish 10.78, and for German 9.13. The first language we clustered was English, were we found that linking devices distorted our results. For instance, for the verb-preposition "make-of" the by far most frequent noun was "course", an undesirable result based on the expression "of course". Thus, we excluded linkers for the English clustering data. For Swedish this problem was less prominent, perhaps due to the use of a different tagger and parser, and in the German data this problem was not observed.

### 5.2 Semantic Role Labeling Results

As exemplified in Table 2, semantic role labeling provided the basis for choosing the best parameter settings for the clustering algorithm. The results for the best setting combination knn, PPMI, and $k$ size 300 with normalized clustering by Ng et al. [28] for all languages are presented in Table 3, which also shows the absolute frequency of image-schematic clusters. We analyzed all clusters, spatial and non-spatial, for their image-schematic content for the chosen 300 clusters. In English, the majority of detected image schema clusters were also labeled with spatial semantic roles, as is the case for German. In Swedish, however, 34% of all clusters were detected in non-spatial clusters. We attribute this to the lower accuracy of the semantic role labeler, especially since a substantially higher number of clusters (66%) were not assigned with a spatial label.

**Table 3.** Total number of clusters per label and number of image schema (IS) clusters

| | English | | Swedish | | German | |
|---|---|---|---|---|---|---|
| Label | Clusters | IS | Clusters | IS | Clusters | IS |
| Spatial | 92 | 74 | 64 | 59 | 88 | 80 |
| Mixed | 49 | 18 | 38 | 22 | 18 | 13 |
| Other | 159 | 18 | 198 | 42 | 194 | 10 |
| Total | 300 | 110 | 300 | 123 | 300 | 103 |

### 5.3 Image Schema Identification Results

Our assumption was that the proposed method groups verb-preposition pairs based on their respective nouns into spatial and potentially image-schematic clusters. Thus, we are interested in how many of the spatial clusters actually are image-schematic. To calculate the accuracy of our method, we compare the number of obtained spatial clusters to the number of image-schematic clusters with a spatial label in relation to the total number of image-schematic clusters, the data of which are presented in Table 3 and 4. For English, this provides 74 image schema clusters in 92 spatial clusters with a total of 110 image schema clusters across the whole set of 300, which provides an accuracy of 80.43% (F-measure: 73.27%). For Swedish, the accuracy is 82.81% (F-measure: 56.38%) which also has the largest number of image schema clusters in the set of non-spatial clusters compared to the other languages, hence the low F-measure. For German, the accuracy is 90.91% (F-measure: 83.77%) because most image schema clusters have a spatial role label. This can be attributed to the manual annotation of the semantic roles performed only in German.

In English the identification of image-schematic structures was conducted by two experts with an inter-annotator 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. For Swedish and German we only had one expert annotate each language for this first experiment.

**Table 4.** Detected image schema clusters

| Image Schemas | English | | Swedish | | German | |
|---|---|---|---|---|---|---|
| Absolute (A) & % | A | % | A | % | A | % |
| CONTAINMENT | 61 | 55 | 81 | 66 | 57 | 55 |
| SOURCE_PATH_GOAL | 22 | 20 | 14 | 11 | 27 | 26 |
| SUPPORT | 16 | 14 | 21 | 17 | 6 | 6 |
| SURFACE | 4 | 4 | 0 | - | 7 | 7 |
| VERTICALITY | 3 | 3 | 3 | 2 | 0 | - |
| Other | 4 | 4 | 4 | 3 | 6 | 6 |
| Total Schemas | **110** | | **123** | | **103** | |

The most commonly identified image schema in all languages is CONTAINMENT as illustrated in Table 4. One of the reasons for this is the description provided by Johnson [13] with the 'inside-border-outside' relationship that establishes a rather general reference which fits many scenarios. To account for the spatial relationships of the image schemas, as well as their conceptual correspondence, it would provide a more accurate account to divide this image schema into a family (and idea supported by e.g. [33,3,12]). While the members of a CONTAINMENT family need to be properly established it is clear that spatial relations such as "being on the outside/inside", "degrees of parthood", "going in/out" and "going through" are fundamentally different despite all belonging to CONTAINMENT. To account for these differences in our data, we would have to conduct a more detailed analysis considering the context for each verb-preposition-noun combination, which could be interesting for further investigations.

Swedish has a higher number of CONTAINMENT since its overall count of verb-preposition pairs is higher than in the other two languages. The second most frequent image schema is SOURCE_PATH_GOAL, such as "fortsätta-på-väg" (continue-along-road), with a lower frequency in Swedish. In contrast, Swedish had more SUPPORT schemas especially in comparison to German, e.g. the German expression "lasten-auf-schultern" (rest-upon-shoulders). 'Other' in the image schema column in Table 1 refers to the collected occurrences of the image schema structures NEAR-FAR, SPLITTING, PART-WHOLE, SCALING and CENTER-PERIPHERY. Other than for SUPPORT the numbers and types of image schemas are quite comparable across languages.

We would like to point out that the numbers presented in Table 3 and 4 are image schema clusters containing several verb-preposition pairs with several feature nouns, e.g. for English we obtain 2.567 image-schematic verb-preposition-noun triples from the 110 clusters. Table 5 provides one CONTAINMENT cluster for each language to exemplify the results. While they are not aligned automatically by our method, not identical and vary in size, they have several features in common, such as all of them referring to the CONTAINER "hands" as in "play-into-hands" or "lie-in-hands".

The resulting repository can then be either represented as those verb-prepositions clusters with the annotated image schema as in Table 5 or in a more detailed way combining pairs with their feature nouns as depicted in Table 6.

**Table 5.** Cluster example CONTAINMENT

| | |
|---|---|
| English: | 'concentrated-in','lie-in', 'lies-in', 'play-into', 'played-into', 'playing-into', 'plays-into', 'sit-on', 'suffer-at', 'suffered-at' |
| Swedish: | 'finnas-i', 'föra-till', 'föras-till', 'förs-till', 'noteras-i' |
| German: | 'fallen-in', 'fällt-in', 'fällt-unter', 'gelangt-in', 'herausgenommen-aus', 'verbleiben-in' |

**Table 6.** Repository example with annotations

| Verb-Prep-Noun | Image Schema | Role Label | Sentence |
|---|---|---|---|
| concentrated-in-areas | CONTAINMENT | Location | ... unemployment is quite concentrated in particular areas... |
| plays-into-hands | CONTAINMENT | Destination | ... which plays straight into the hands of the radicals... |
| continue-along-road | SOURCE_PATH_GOAL | Direction | ... must be encouraged to continue along this road ... |

## 6 Discussion

In our corpus the manifestations of image schemas in natural language related to both the abstract and the concrete. For instance, in "*Beziehungen* mit neuem Leben erfüllen" (fill *relationships* with new life) "relationships" are abstract CONTAINERs while in "sie füllen ihre *Taschen* mit Gold" (they are filling their *pockets* with gold) "pockets" (real or hypothetical) are concrete CONTAINERs. This juxtaposition of abstract and concrete concepts makes the corpus a good dataset for investigations into the nature of image schemas since they offer to ground abstract phenomena in physical sensorimotor experiences. In addition, this analogous relationship between the concrete and the abstract assists the task of annotating the clusters for image schemas, for if "pockets" is a container then "relationships" has to be evaluated against the CONTAINER criteria as well.

The most challenging part of our method was the semantic role labeling of preposition senses, which had to be approached with different methods for the three different languages in our study. In fact, we believe that the low F-measure in Swedish might be attributed to the role labeling, since both the English high-quality labels and the German manual annotation of semantic roles returned more satisfactory F-measures. Naturally it would be preferable if similar tools for semantic labeling were available for the investigated languages. One possible solution would be to follow the method in [36] and from the beginning manually label the clusters. The consequences this might have had on our results are of less importance than for studies that aim at providing an in-depth crosslingual comparison of image schemas, which we intend to do as future work. One solution could be to crowdsource the spatial role labeling task.

Another important aspect in need of improvement is the manual annotation of image schemas. It is error-prone and biased by the human evaluators. This problem was highlighted through a preliminary crosslingual comparison. For instance, "put-on-market" was annotated as SUPPORT in English, CONTAINMENT in German, and as none in

Swedish. This disjoint annotation is partly due to individual annotators, but also due to the different connotations natural languages have and might therefore not be wrong in itself but provides little confidence in terms of automated cross-lingual comparisons. To improve the annotation process, we hope to rely on different supervised approaches based on good examples from the natural language image schema repository we created for this paper. Alternatively, we might consider crowdsourcing also for this step.

The current analysis extracted image schemas from natural languages pertaining to the same language family. To really confirm that our method can be used effectively to extract image schemas from multilingual corpora, we need to test it on other language families as well, a process that has been started but not finished in time for this paper due to the time-consuming annotation process.

Regarding the results, the spatial clusters return an overwhelming numbers of CONTAINMENT schemas. While CONTAINMENT undeniably is one of the most essential image schemas, there is room for improvement here. As previously observed [33,12], image schemas do not always appear in isolation, but rather as families. During the process of annotating the image schemas, many kinds of CONTAINMENT schemas could be detected. Following previous approaches [3,8], it would be interesting to analyze the elements involved in image schemas (e.g. border, inside, outside of CONTAINMENT) and their interaction in natural language instead of identifying abstract image schemas only.

Our results show that findings from investigations into spatial language can effectively be used to extract image schemas from natural languages. They also show that image schemas are prevalent in natural language, even highly abstract language. Knowing that the abstract "relationships" from the above example have the same underlying image schema of CONTAINMENT as the physical "pockets" can further our understanding of the influence of sensorimotor experiences on language use. It means that we conceptualize both in certain contexts as containers with a fill level. Our study found a rather similar distribution of image schemas across three languages. The resulting repository of multilingual expressions annotated with image schemas provides a good starting point for a crosslingual comparison, which we intend to do including other language families than considered here. Such investigation can contribute to research on the universality of image schemas.

## 7 Conclusion and future work

We present a method to semi-automatically extract image schemas from natural language, which provides promising results that confirm our assumption that verb-preposition pairs with their context nouns as features are good indicators of spatial and also image-schematic language. We exemplify the approach in English, Swedish, and German.

Parts of the method are manual and still in a preliminary stage. Future work therefore includes to use the results from this study as examples for a supervised approach to work towards a method requiring less manual effort. We also intend to broaden the current experiment to different language families and test with more feature combinations than the three word classes used herein.

## References


1. Alomari, M., Duckworth, P., Hogg, D., Cohn, A.: Learning of object properties, spatial relations, and actions for embodied agents from language and vision. In: To be confirmed. AAAI Press (2017)
2. Baroni, M., Dinu, G., Kruszewski, G.: Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In: ACL (1). pp. 238–247 (2014)
3. Bennett, B., Cialone, C.: Corpus guided sense cluster analysis: a methodology for ontology development (with examples from the spatial domain). In: Garbacz, P., Kutz, O. (eds.) 8th International Conference on Formal Ontology in Information Systems (FOIS). Frontiers in Artificial Intelligence and Applications, vol. 267, pp. 213–226. IOS Press (2014)
4. Björkelund, A., Hafdell, L., Nugues, P.: Multilingual semantic role labeling. In: Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task. pp. 43–48. Association for Computational Linguistics (2009)
5. Chen, D., Manning, C.D.: A fast and accurate dependency parser using neural networks. In: EMNLP. pp. 740–750 (2014)
6. Dodge, E., Lakoff, G.: Image schemas: From linguistic analysis to neural grounding. In: Hampe, B., Grady, J.E. (eds.) From perception to meaning: Image schemas in cognitive linguistics, pp. 57–91. Mouton de Gruyter, Berlin (2005)
7. Gibbs, R.W., Colston, H.L.: The cognitive psychological reality of image schemas and their transformation. Cognitive Linguistics 6, 347–378 (1995)
8. Gromann, D., Hedblom, M.M.: Breaking down finance: A method for concept simplification by identifying movement structures from the image schema path-following. In: Proc. of the Joint Ontology Workshops (JOWO) (2016)
9. Gromann, D., Hedblom, M.M.: Kinesthetic mind reader: A method to identify image schemas in natural language. In: Proceedings of Advancements in Cognitive Systems (2017)
10. Guerin, F.: Learning like a baby: A survey of AI approaches. The Knowledge Engineering Review 00(0), 1–22 (2008)
11. Hall, J., Nivre, J., Nilsson, J.: A hybrid constituency-dependency parser for swedish. In: Proceedings of NODALIDA. pp. 284–287 (2007)
12. Hedblom, M.M., Kutz, O., Neuhaus, F.: Choosing the Right Path: Image Schema Theory as a Foundation for Concept Invention. Journal of Artificial General Intelligence 6(1), 22–54 (2015)
13. Johnson, M.: The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reasoning. The University of Chicago Press (1987)
14. Kingsbury, P., Palmer, M.: From treebank to propbank. In: LREC. pp. 1989–1993. Citeseer (2002)
15. Koehn, P.: Europarl: A parallel corpus for statistical machine translation. In: MT summit. vol. 5, pp. 79–86 (2005)
16. Kollar, T., Tellex, S., Roy, D., Roy, N.: Grounding verbs of motion in natural language commands to robots. In: Experimental robotics. pp. 31–47. Springer (2014)
17. Kordjamshidi, P., Van Otterlo, M., Moens, M.F.: Spatial role labeling: Towards extraction of spatial relations from natural language. ACM Transactions on Speech and Language Processing (TSLP) 8(3), 4 (2011)
18. Kövecses, Z.: Metaphor: A Practical Introduction. Oxford University Press, USA (2010)
19. Krishnamurthy, J., Kollar, T.: Jointly learning to parse and perceive: Connecting natural language to the physical world. Transactions of the Association for Computational Linguistics 1, 193–206 (2013)
20. Lakoff, G.: Women, fire, and dangerous things. what categories reveal about the mind. The University of Chicago Press (1987)

21. Lakoff, G., Johnson, M.: Metaphors We Live By. University of Chicago Press (1980)
22. Litkowski, K.: Pattern Dictionary of English Prepositions. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. vol. 1, pp. 1274–1283. Baltimore, Maryland (2014)
23. Mandler, J.M.: How to build a baby: Ii. conceptual primitives. Psychological review 99(4), 587 (1992)
24. Mandler, J.M., Pagán Cánovas, C.: On defining image schemas. Language and Cognition pp. 1–23 (2014)
25. Mani, I., Pustejovsky, J.: Interpreting motion: Grounded representations for spatial language. No. 5 in Explorations in Language and Space, Oxford University Press (2012)
26. Mason, Z.J.: Cormet: a computational, corpus-based conventional metaphor extraction system. Computational linguistics 30(1), 23–44 (2004)
27. Misra, D.K., Sung, J., Lee, K., Saxena, A.: Tell me dave: Context-sensitive grounding of natural language to manipulation instructions. The International Journal of Robotics Research 35(1-3), 281–300 (2016)
28. Ng, A.Y., Jordan, M.I., Weiss, Y., et al.: On spectral clustering: Analysis and an algorithm. In: NIPS. vol. 14, pp. 849–856 (2001)
29. Östling, R.: Stagger: An open-source part of speech tagger for swedish. Northern European Journal of Language Technology (NEJLT) 3, 1–18 (2013)
30. Pauwels, P.: Levels of metaphorization: The case of put. In: Goossens, L. (ed.) By Word of Mouth: Metaphor, metonymy and linguistic action in a cognitive perspective, pp. 125–158. John Benjamins Publishing Company, Amsterdam (1995)
31. Punyakanok, V., Roth, D., Yih, W.: The importance of syntactic parsing and inference in semantic role labeling. Computational Linguistics 34(2) (2008)
32. Roth, M., Lapata, M.: Neural semantic role labeling with dependency path embeddings. arXiv preprint arXiv:1605.07515 (2016)
33. Santibáñez, F.: The object image-schema and other dependent schemas. Atlantis 24(2), 183–201 (2002)
34. Shapiro, L.: Embodied cognition. New problems of philosophy, Routledge, London and New York (2011)
35. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on pattern analysis and machine intelligence 22(8), 888–905 (2000)
36. Shutova, E., Sun, L., Gutierrez, D., Lichtenstein, P., Narayanan, S.: Multilingual metaphor processing: Experiments with semi-supervised and unsupervised learning. Computational Linguistics (2016), forthcoming
37. Spranger, M.: The evolution of grounded spatial language (2016)
38. Sun, L., Korhonen, A.: Improving verb clustering with automatically acquired selectional preferences. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2. pp. 638–647. Association for Computational Linguistics (2009)
39. Talmy, L.: The fundamental system of spatial schemas in language. In: Hampe, B., Grady, J.E. (eds.) From perception to meaning: Image schemas in cognitive linguistics, Cognitive Linguistics Research, vol. 29, pp. 199–234. Walter de Gruyter (2005)
40. Von Luxburg, U.: A tutorial on spectral clustering. Statistics and computing 17(4), 395–416 (2007)
41. Xu, Z., Ke, Y.: Effective and efficient spectral clustering on text and link data. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. pp. 357–366. ACM (2016)
42. Zlatev, J.: Spatial semantics. In: Geeraerts, D., Cuyckens, H. (eds.) The Oxford Handbook of Cognitive Linguistics, pp. 318–350. Oxford University Press (2010)