A Grounded Cognitive Model for Metaphor Acquisition

Sushobhan Nayak and **Amitabha Mukerjee**
Indian Institute of Technology, Kanpur
India
{snayak, amit}@iitk.ac.in

Abstract

Metaphors are central to our language and thought process, and modelling them computationally is imperative for reproducing human cognitive abilities. In this work, we propose a plausible grounded cognitive model for artificial metaphor acquisition. We put forward a rule-based metaphor acquisition system, which doesn't make use of any prior 'seed metaphor set'. Through correlation between a video and co-occurring commentaries, we show that these rules can be acquired in an unsupervised manner by an early learner capable of abstracting from multi-modal sensory input. From these grounded linguistic concepts, we derive classes based on lexico-syntactical language properties. Based on the selectional preferences of these linguistic elements, metaphorical mappings between source and target domains are acquired.

Introduction

The seminal work by (Lakoff and Johnson 1980) changed our understanding of and attitude towards metaphors. Metaphors, which had till then been considered a 'device of the poetic imagination and the rhetoric flourish', soon came to be regarded as an integral part of our action and thought processes: "Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature"(Lakoff and Johnson 1980). Consider the following:

- You are *wasting* my time. (TIME IS MONEY)
- We need to *combat* inflation. (INFLATION IS AN ENTITY)

Such usages are so common in a conversation that we fail to recognize that these are abstract metaphors derived from concrete concepts. Arguably, metaphors have a central role in human cognition and thought, and it would be impossible for an artificial system to reproduce human-level intelligence without some provision for metaphor acquisition and interpretation.

The work on metaphors in NLP may be characterized in terms of three phases. Earlier *rule-based* attempts such as (Wilks 1975; Fass 1991) were based on hand-coded knowledge and metaphors were identified as a violation of selectional restrictions in a given context (e.g. "my car drinks gasoline").

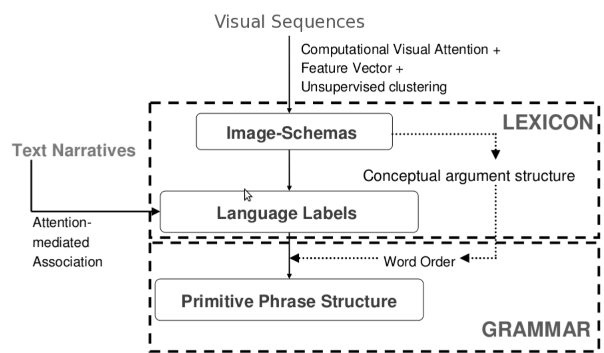

Figure 1: *Acquiring a grounded language*. Metaphor discovery relies on similarity in sensorimotor space. Our approach discovers image-schemas from perceptual input unaided by language, and then associates these unsupervised notions with units from text narratives. It then uses these semantic associations to map metaphoric usage.

Other models have attempted to use syntactic and co-occurrence data across large corpora to identify metaphors. We may call these attempts as *corpus-driven*; work here may include (Shutova, Sun, and Korhonen 2010) who demonstrates metaphor paraphrasing using noun-verb clustering, or (Kintsch 2000) who effectively uses Latent Semantic Analysis to interpret metaphors like "My lawyer is a shark". Cormet (Mason 2004) is able to find mappings given separated datasets for two domains, e.g. it finds LIQUID → MONEY once provided with LAB and FINANCE specific corpora to train from.

Corpus-based approaches keep the metaphor mapping implicit, i.e. while the system can produce many metaphorical sentences statistically, it has no underlying representation for mappings such as LIFE IS A CONTAINER (for example, if it is presented with 'stir excitement', 'throw remark' etc., it identifies that 'swallow anger' is a metaphor, but it never explicitly acquires the concept of ANGER IS AN OB-

JECT). Furthermore, they often need a hand-coded 'seed set' or metaphor repository to do further learning. Also, due to their sole reliance on verbs to find relationship between concepts, they fail to acquire ontological metaphors like CONTAINER metaphors that are heavily preposition dependent. While these works may be able to identify metaphor use, they are far from the holy grail for AI and cognitive science, which is to reproduce human cognitive abilities in computational artefacts. Their coverage is broad, they are shallow in depth.

These models treat metaphors as statistically rare structures in text, and overlook the fact that metaphors operate on a rich semantics linked to sensorimotor models. The similarity underlying metaphors highlights a single aspect of meaning in a way that formal symbols systems are not designed to focus on. A third category of work, which we may call *embodied modeling*, attempts to learn such mappings. This can be seen in cognitively motivated works such as (Narayanan 1997; 1999), who build on the x-schema model for (Bailey 1997) pre-motor cortical representation of movement, and extends this to consider possible metaphoric usage such as "India releases the stranglehold on business".

The embodied approach is intellectually appealing and elegant, but it is hard to scale up due to two reasons. First, it is enormously difficult to posit new semantic structures for the vast array of linguistic concepts. Secondly, even given the semantics, it is difficult to relate it to language without some knowledge of syntactic structure and how it relates to this semantics. In this work, we attempt to present a grounded models where the rich semantic models underlying such metaphorical transfers need not be carefully engineered, but may be learned based on simple sensori-motor inputs ((Roy, Hsiao, and Mavridis 2004)). In our approach, we consider perceptual inputs in terms of image sequences and discover a set of sensorimotor object, action and relation clusters, and then associate these with words from text (Fig 1). The action and relation clusters are image schemas and a cluster such as "in" would also identify objects that are likely to be containers. Next, the syntactic structures are discovered (in a very primitive form) from this set of narratives, and structures such as "the big square chases the circle, little square" or "the circle, square is, are in the box" emerge. This constitute the Grammar aspect in Fig 1. Since this process is completely unsupervised, it would be possible to scale it up by applying it to novel situations. To demonstrate the generalness of applying such grounded linguistic knowledge to general text, we search for structures similar to what we have discovered in the Brown corpus, and identify several instances of metaphorical usage related to the "in the X" construction. Thus, this entirely process, starting from a multimodal input, is able to discover perceptual spatio-temporal pattern, map them to linguistic units, relate these to a primitive notion of syntax, and exploit this conceptual basis to acquire metaphorical mappings in a natural way that emulates language learning in an early learner.

A Semantics-First Manifesto

How can a system *acquire* metaphors - i.e. learn that certain concepts may be mapped from a source domain to a target domain based on a particular facet of similarity? The models discussed so far, while being excellent pieces of NLP, do not address this question, assuming somehow that metaphors already exist in the system. However, (Lakoff and Johnson 1980) suggest that we should look at metaphor acquisition the same way we look at language acquisition, and not something that is interpreted/acquired *after* we have acquired the basic nuances of language. Just as language is a representation of our thought process, so are metaphors. Thus, it is critical to treat metaphors in the same sort of grounded view which is applied to language as a whole. The following two observations motivate towards our model:

Observation 1: Initial Grounding

There is ample evidence in literature to suggest that basic linguistic forms may be grounded(Roy and Reiter 2005). The human infant acquires words from a grounded context, and forms perceptual schemas which reflect concepts that have arisen pre-linguistically, and are therefore independent of the words used. These schemas are often fuzzy and although they may be approximated by discrete propositional structures, the grounded perceptual schema remains available as a fallback for disambiguating conflicts. Such models, often called *Image Schema* in Cognitive Linguistics(Langacker 1987) or *Perceptual Schema* in Experimental Psychology (Mandler 2004), involve abstractions on low-level features extracted from sensory-motor modalities. It is widely believed in cognitive science that the process of category formation operates on these inputs to define the structure of our conceptual space.

Observation 2: Extension Through Language

Only a small initial set of grounded linguistic concepts is needed. The majority of one's vocabulary is learned later purely from linguistic inputs(Bloom 2000). Even metaphors are most likely learned from language usage only, without any physical grounding. The process by which this is achieved is not dealt with in this work, but it would seem that a similar physical experience can lead to different metaphorical mappings in different cultures. Given a large corpus that an early learner is being exposed to, it may learn that certain words share some co-occurrence statistics, forming natural clusters. As a quick example, searching for words that occur in similar contexts to 'love' in texts *Moby Dick* by Melville (using `text.similar()` [1] function in Python NLTK Toolkit), gives the following outputs:

```
Man Sea Ahab Air Bone Captain Chase
Death Fear Hope Land
```

Given such strong evidence for grouping of Objects/Entities (Man, Ahab, Captain) with emotions (love, fear, hope), an early learner, based on such word usage, along with an internal grounding for the Entity/Object class, may begin to impart object properties on feelings, leading to the glimmerings of something like the FEELINGS ARE OBJECTS metaphor.

Language is faced with the difficult task of having to handle the infinite variation of the world in terms of a much

[1]This function takes a word w, finds all contexts w_1ww_2, then finds all words w' that appear in the same context, i.e. $w_1w'w_2$

smaller number of constructions. A key process in language is to apply the same linguistic structures to different situations based on similarity. This is the primary argument in this work, which suggests that pre-linguistic concepts such as object categorization, containment, etc. - are used as the child tries to ground linguistic elements in terms of already acquired perceptual schemas. Therefore, nouns describing physical objects/substances seem to be easier to ground, while abstract concepts are slower to acquire. Through exposure to common conversation, the child subsequently learns syntactic features/grammar of the language(Siskind 1996). Consider an agent with this much information and competence. When the agent, by virtue of its knowledge of grammar, comes across sentences in which words describing emotions appear in the same context as a grounded object, it's but natural on a such a limited first evidence to impart grounded aspects of an object to this abstract concept. Later on, through more exposure, the agent might imbibe the abstract nature of the linguistic element.

To determine how far language usage alone can help shape the concept of metaphors, we compiled a list of sentences from (Lakoff and Johnson 1980) and (Lakoff, Espenson, and Schwartz 1991) that correspond to the ontological metaphor-mappings for Containers, Objects and Substances. Ontological metaphors are "ways of viewing events, activities, emotions, ideas, etc., as entities and substances"(Lakoff and Johnson 1980). Our experiences with physical objects (especially our own bodies) provide the basis for an extraordinarily wide variety of ontological metaphors. The salient findings of this search were:

- Of the 85 sentences denoting Container metaphors, in 65, the abstract idea was imparted the image schema of a container based only on the prepositions *in/out*. In the rest 20, adjectives (*full, empty*) and verbs (*explode, erupt, fill*) took up the mantle.
- In all of the 63 sentences for Object metaphor, the Object property was imparted to the concept because VERB(A,B) took object arguments, i.e. verbs were the primary basis of metaphor mapping.
- Of the 42 sentences for Substance metaphors, 17 mappings were done based on adjectives (*more, less*) while the rest were of the type *Container contains Substance*, i.e. first the Container property was imparted, and then whatever was supposed to be inside the container was called a substance.

We observe that nouns, verbs and prepositions play a pivotal role in metaphorical mappings. But syntactic information alone may not be adequate in discovering rule-based metaphorical mappings; the possibility of over-generation is there at every step. Thus, for these mappings to reflect aspects of the human thought process, they would need to be grounded. The question we set out to address is: *What would a primitive system need to be able of discover such mappings by itself?* In the following section, using a simple video and co-occurring commentaries, we show that this discovery, indeed, may be possible. While the grounding is restricted to a very small domain, we assume that similar grounding is possible in other domains to generalize the argument for a large corpus.

In the system presented next, we proceed in two stages. In the first (see Fig 1), a perceptual agent is exposed to a simple geometric video(from (Heider and Simmel 1944). See Figure 2), and acquires concepts of objects, actions, and relations based on shape constancy and the relative motions of the objects. Later, the learner is exposed to a series of commentaries[2], where it is able to acquire a set of nouns, verbs and prepositions that relate to these classes of concepts. At this point, the system has a set of concepts which are very primitive (it is known only in the domain of this single video) and it knows the maps for these concepts to some linguistic labels. Each label-semantics pair may be thought of as a primitive form of *symbolization* in Cognitive Grammar (Langacker 1987). This is the first phase of the learning which corresponds to the *initial grounding* of the first observation above.

In the next phase, the system is exposed to broader linguistic corpus, and it encounters these symbolic units along with others for which also it is assumed to have similar primitive structures. By correlating among these structures based on the initial grounding that it has acquired earlier, it is able to enrich the primitive semantics of these units. By analysing the linguistic co-occurrences, it is able to learn selectional preferences from the corpus. We show how these selectional preferences lead to acquisition of clusters that suggest some of the metaphorical structures named by (Lakoff and Johnson 1980). In this way, we attempt to demonstrate, albeit in a much weaker form than the initial stages, the mechanisms of *Extension through language* (Observation 2).

Grounding Linguistic Concepts

Almost all types of ontological metaphors come under three broad categories of an **Object**, a **Substance** or a **Container**. In fact, these concepts emerge directly for an early learner through physical experience. "We experience ourselves as entities, separate from the rest of the world - as containers with an inside and outside. ... We experience ourselves as being made of substances-..."(Lakoff and Johnson 1980). Abstract concepts like `courage`, based on context, act like CONTAINER/OBJECT/SUBSTANCE etc., which are physical concepts, so that, in a way, abstract concepts are also characterized in terms of image schemas. In this section, using the aforementioned video(Fig 2) and commentaries, we try to ground concepts of objects(nouns), verbs and containment(Fig 1).

Noun/Entity Acquisition

We follow (Ballard and Yu 2003) and (Mukerjee and Sarkar 2007) to acquire nouns representing salient objects(squares and the circle in the video) from sentential input. Attentive focus is used to constrain the region of visual salience in

[2]A linguistic database consisting of a co-occurring narrative with 36 descriptions of the video which exhibit a wide range of linguistic variation both in focus (perspective) and on lexical and construction choice.

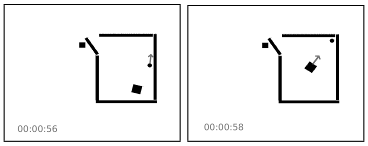

Figure 2: *Multimodal input: 2D video "Chase"*: Three easily segmented shapes, [big-square], [small-square] and [circle] interact playfully (velocities shown with arrows).

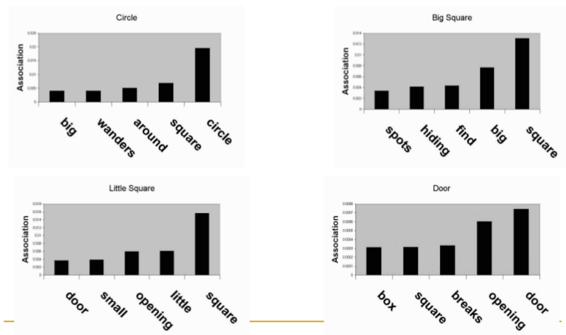

Figure 3: Object label associations: High association value is evidence that `square`, `circle`, `door` have been mapped from perceptual schemata to language domain.

each frame of the video, and identify the constituents participating in an action. To identify the nouns describing the participants in the relation, we use word-object association. The word-object association is estimated using the product of mutual information of word w_i and object o_j with their joint probability.

$$A = Pr(w_i, o_j) \log \frac{Pr(w_i, o_j)}{Pr(w_i)Pr(o_j)}$$

We calculate the product of joint probability and mutual information because if $W (= \cup_i w_i)$ and $O (= \cup_i o_i)$ are two random variables then their Mutual Information $I(W, O)$ would be

$$I(W, O) = \sum_i \sum_j Pr(w_i, o_j) \log \frac{Pr(w_i, o_j)}{Pr(w_i)Pr(o_j)}$$

and $Pr(w_i, o_j) \log \frac{Pr(w_i, o_j)}{Pr(w_i)Pr(o_j)}$ would be the contribution of each word object pair.

Nouns `square`, `circle`, `door` are learned by the system (Fig 3) and hence are distinguished from a previously unintelligible set of words in the discourse.

Verb Acquisition

In a previous work(Mukerjee, Neema, and Nayak 2011), we use motion features to find VERB(Subject, Object) mappings for coreference resolution. We consider two-agent spatial interactions, which correspond to verbs with two arguments. The system considers pairs of objects attended to within a short timespan, and computes two inner-product features a) *pos·velDiff* $[(\vec{x}_B - \vec{x}_A) \cdot (\vec{v}_B - \vec{v}_A)]$ and b) *pos·velSum* $[(\vec{x}_B - \vec{x}_A) \cdot (\vec{v}_B + \vec{v}_A)]$. The temporal histories of these feature vectors are then clustered using the temporal mining algorithm *Merge Neural Gas* . Four action clusters are discovered, two of which correspond to **[come-closer]** and **[move-away]**, and two correspond to **[chase]**(Figure 4). Chase has two clusters because it is asymmetric, and the primary attention may be on the *chaser* (Cluster 3) or on the *chased* (Cluster 4). By computing the feature vectors with the referents switched, the system can, by itself, determine this alternation. In the noun acquisition stage, we have already extracted entities `square`, `circle` and `door`. We exclude them and the most frequent words like `the` from the commentaries. We correlate 1-grams from the commentaries with each cluster to find the strongest association for the action clusters with the utterances. From the results, CHASE() verb shows a high degree of association with two clusters. These clusters are therefore taken as an representation of the image schema of CHASE() for this artificial system.

CLUSTER 1 (Come-Close)		CLUSTER 2 (Move-Away)		CLUSTER 3 (Chase)		CLUSTER 4 (Chase)	
ONE WORD LONG LINGUISTIC LABELS(MONOGRAMS)							
corner	0.077	away	0.069	chase	0.671	chase	0.429
move	0.055	move	0.055	other	0.185	after	0.112
attack	0.042	chase	0.049	around	0.183	out	0.033

Figure 4: Table showing the association of the motion clusters with monograms. Notice that Clusters 3 and 4 show a close association with verb `chase`.

Spatial Preposition Acquisition

We have investigated two methods of spatial preposition acquisition task elsewhere(Mukerjee and Sarkar 2007; Nayak and Mukerjee 2012). Here, we detail the work in (Mukerjee and Sarkar 2007). *Stolen Voronoi area* features are used for clustering using a self-organizing map (Kohonen) map. These spatial relation clusters are then matched with words occurring in the user commentaries, using the same mutual information measure as used for nouns. Since containment structures involve two objects, phrases containing two nouns/object names are considered and the feature vector relating the two objects is calculated and checked for the cluster they lie in. Similarly cluster labels for each phrase are collected and automatas are constructed to represent the transition between clusters for each phrase (e.g. circle coming out of the room will transition from one state **[in]** to another **[out]**). These automatas were then associated with their corresponding phrases. Strong associations emerge between the words `out` and `in` (see the following table).

1-Grams	Association	1-Grams	Association
breaks	0.16	to	0.18
out	**0.22**	goes	0.18
of	**0.22**	**in**	**0.25**

The left part of the table is association of words when the transition is from cluster **[in]** to cluster **[out]**. The right-hand side of the table is for the reverse transition. `in` and `out` emerge as monograms having strong correlation with the perceptual spatial schema transition. `out` is almost al-ways uttered as `out of` in the commentaries, leading to `of` being salient too.

To summarize, we set out to investigate if an artificial system, with limited sensory input, would be able to distinguish linguistic concepts like nouns, verbs and prepositions. We showed, drawing findings from various previous works, that this indeed may be possible. We also saw that, the notion of trajectors(circle, square) executing containment can be learned. Associating ‘container’ with IN schema has been described in detail in (Nayak and Mukerjee 2012). It has also the schema for the verb CHASE() at it’s disposal. And it also can distinguish prepositions IN and OUT and containment. We will assume that these are the grounded linguistic forms that the system is capable of handling. For demonstrative purposes, we also assume that the system has a grounded form for SPEND(). With this world view, we will now try to acquire metaphorical mappings for this system through corpus manipulation. We next show that this notion of trajector objects and containers can be extracted from syntactic information alone, once their grounded forms have been acquired, so that abstract concepts that are syntactically grouped with these grounded forms imbibe the grounded properties.

Finding Metaphorical Maps

In Observation 2, we discussed the importance of acquiring metaphor maps from language usage. With a few basic grounded concepts, we can learn myriads of similar linguistic concepts through their lexical and syntactic similarity. Consider the following examples:

- You’re wasting my time.
- How do you spend your time these days?

These sentences would still feel natural if one replaces `time` with `money`. In fact, since the idea TIME IS MONEY is ingrained in our thought process, these concepts are used inter-changeably in forming sentences. It’s therefore prudent to assume that metaphorical concepts would occur in similar lexico-syntactic environments in language usage. Of course, to recognize such syntactic similarities, a system needs to have some underlying grammar that it can follow. Specifically, we would need argument structure rules that can convert ‘wasting my time’ to the functional form WASTE(`time`) so that association between verbs and their predicates can be calculated. At present, our system is not privy to any grammar implementation. However, similar attempts can be found in literature; for example, Siskind(Siskind 1996) presents cross-situational techniques for learning word-to-meaning mappings. For the present work, in the following section, we show a plausible way in which such argument structure can be extracted.

Discovering Argument Structure

To emulate the syntactic structure acquisition task, we used ADIOS(Solan et al. 2002), which finds syntactic categories from a corpus without requiring pre-judging of either the scope of the primitives or of their classification, in an unsupervised way that is cognitively simple and plausible for a child. It first creates a Representational Data Structure(RDS) by morphologically segmenting the input sentences and creating directed edges between vertices corresponding to transitions in the corpus. It then repeatedly scans and modifies the RDS to detect significant patterns. The question we are trying to address is, “*Given the grounded concepts, can a primitive agent discover/distinguish them from a stream of unconstrained words? Also, can it discover linguistic units belonging to similar classes as the grounded concepts?*” Results show that this might, indeed, be achievable, albeit in a primitive way. Two of the outcomes of the test run of ADIOS on our corpus are presented below(all the elements inside a square bracket belong to the same equivalent class and are exchangeable in a sentence):

$$\left[[circle] \begin{bmatrix} is \\ went \\ goes \end{bmatrix} \right] \begin{bmatrix} in & inside & into \\ at & by & \end{bmatrix}$$

$$\left[\left[the \begin{bmatrix} ball \\ door \\ box \\ square \end{bmatrix} \right] [circle] \right] \begin{bmatrix} move \\ came \\ got \end{bmatrix} into$$

Notice that the spatial prepositions/descriptors have been grouped as one class in both constructions. Also, there is demarcation between the objects(`circle`) and verbs(`is`, `went`). While the objects-descriptors have been segregated from unfamiliar linguistic elements, similar unforeseen elements(`box`, `at` etc.) have been grouped in respective equivalence classes, thereby imparting properties of the known elements to the unknown ones (`box` as the agent). Further experiments with a detailed corpus (Brown) provides us with the following structures:

$$in\ the \begin{bmatrix} building & war & fight & car \\ group & death & woods & cellar \end{bmatrix},$$

$$the \begin{bmatrix} other & second \\ first & last \end{bmatrix} \begin{bmatrix} things & action \\ noise & step \end{bmatrix} in\ the$$

The power of this simple method is that it is able to distinguish `the`, `last` etc. from the prepositional object of IN, in this case `war`, `group`, `cellar` etc. Also notice that it groups seemingly abstract concepts like `war`, `fight` and `death` with physical containers like `car` and `building`, giving the evidence for possible containment metaphorical mappings. It further segregates `things`, `noise` etc. as trajectors, imparting in them the properties of Objects. This gives the preliminary evidence that sentences can be broken down into argument structures from purely statistical syntactic knowledge from the exposed corpus.

Selectional Preference

We saw in Observation 2 that verbs play a major role in imparting metaphorical meanings. The said observation is also supported by (Shutova and Teufel. 2010), who claim that in 164 out of 241 metaphorical sentences, metaphoricity was introduced by verbs. Following discussion in the above paragraph, *given grammatical relations*, an agent should be able to find verbs similar in semantic or syntactic aspects to it's repository of grounded forms.

The metric used the most in literature to measure regularities of a verb w.r.t. the semantic class of its argument (subjects, objects etc.) is **selectional preference (SP)**(Resnik 1993). Some formulations of SP have been used previously for word-sense disambiguation(Resnik 1993) and metaphor interpretation(Mason 2004). While they have only been used for finding verb preferences, we will adapt them to include prepositional preferences too, so that we are able to learn more metaphors, especially containment metaphors, which will be otherwise hard to learn.

We follow the formulation presented in (Resnik 1993). Suppose predicate p selects class c for the syntactic relation r, which we represent as *selects(p,r,c)*. For example, '*drink* takes *LIQUID* at object position' is represented as *selects(drink, object, LIQUID)*. The *selectional association*($A(p,r,c)$) of class c for predicate p is then defined as:

$$S(p,r) = D(P(c|p,r)||P(c)) = \sum_c P(c|p,r) \log \frac{P(c|p,r)}{P(c)}$$

$$A(p,r,c) = \frac{1}{S(p,r)} P(c|p,r) \log \frac{P(c|p,r)}{P(c)}$$

While verbs have different syntactic relations like `verb-object`, `subject-verb` etc., the prepositions we are considering, have only one relation to the trailing noun, that of *Object of Preposition (***pobj***)*(Marie-Catherine de Marneffe and Manning 2006). So, the formulation essentially remains the same and effects of the variable r are nullified.

We use WordNet(Feinerer and Hornik 2011) as a knowledge-base for class c. WordNet was developed as a system that would be consistent with the knowledge acquired over the years about how human beings process language. Since an early learner, like our system here, would not have detailed information of all the synsets of a particular concept, we make use of only the lexical file types (25 in number), which encompass concepts like `quantity` and `possession`. These are the top level abstractions and we assume that an early learner is at a cognitive state where it has notions of these high level concepts.

Finding SPs

We used the Brown Corpus to test our model. All the sentences involving the grounded concepts were extracted. The sentences with prepositions were converted to the functional form of PREP(*pobj*) in a rather simple way: the first occurrence of a singular or mass noun(NN) after the preposition in the tagged corpus was assigned to the concept. For example, the sentence fragment *into a hot cauldron* is converted to INTO(cauldron). [3] Handling sentences pertaining to the verb groups was tricky. The Stanford Parser(Marie-Catherine de Marneffe and Manning 2006) was first used to extract VERB(*object*) relations. But owing to the large number of misclassification of dependencies, the dataset was rechecked by human annotators to correct discrepancies wherever present.

Following Resnik, $P(c|p,r) = freq(p,r,c)/freq(p,r)$ with

$$freq(p,r,c) = \sum_{w \epsilon c} \frac{count(p,r,w)}{classes(w)}$$

where $count(p,r,w)$ is the number of time w occurred, and $classes(w)$ is the number of classes it belongs to. To computationally manage a top-level WordNet semantic node, we approximated it over only those words which are part of the derived corpus.

Discovering Metaphor Mappings

CHASE() is seriously under-represented in Brown(<10 occurrences). For the SPEND() verb class, selectional association for classes `possession` and `time` are 0.269 and 0.731 respectively. The representative nouns, in decreasing order of frequency are:

`time day money dollar week year hour`

This association leads to TIME IS POSSESSION metaphorical mapping. Since *money* is the largest contributor to `possession` in this context, it also leads to TIME IS MONEY metaphor. Also, this example shows why corpus only metaphor acquisition *without* a grounded world view is problematic. In (Mason 2004), the class which has higher selectional association is considered a base class. However, in this case, the reverse happens. Here TIME is the target domain and MONEY is the source domain. Therefore, source or target domain distinction can be done based on a world-view of grounded concept only.

Similarly, when IN() is considered, the following representative nouns and classes are obtained:

`way world order case room place area`
STATE COGNITION ACT COMMUNICATION

Individual selectional association(SA)s for the words and classes together was obtained, the results of which are presented in the following table.

Word/Class	SA	Word/Class	SA
way	0.0968	STATE	0.093
world	0.1028	COGNITION	0.061
order	0.1062	LOCATION	0.08

We notice that, due to these selectional preference mappings on the preposition, these concepts have been associated with containment, leading to the following being acquired:

- STATE/COGNITION AS A CONTAINER
- LOCATION/PLACE AS A CONTAINER

[3] One might notice that this technique for finding the head of the object of a preposition is problematic in the sense that a structure of the form *in iron cauldron* will be taken as *in iron* and not *in cauldron*. However, statistically, less than 1% of occurrences (in Brown corpus) are of the type where the immediate noun following the 'IN' is not associated with it. Furthermore, the Selectional Preference(SP) of 'IN' with 'IRON' over the whole corpus would be meagre, eliminating any effect of a miscalculated prepositional head.

• COUNTRY/LIFE/MIND AS A CONTAINER

Also notice that these are valid ontological metaphors already in vogue in English, and most of them appear in (Lakoff and Johnson 1980).

Conclusion and Future Work

We presented a grounded cognitive model for metaphor acquisition in English language. Without any prior seed set or prior knowledge, we showed that a primitive artificial agent with multi-modal input handling and feature extraction capability could internalize linguistic concepts of noun, verb and prepositions. Through a study of the discourse, we proposed a rule based approach for metaphor acquisition that goes beyond the linguistic only manipulation and integrates metaphorical mappings as a part of the agent's cognition. We show that, the system is capable of finding important Ontological Metaphors.

While being a novel system based on a novel approach, it nonetheless has a few shortcomings. The co-occurring commentaries are inadequate to learn more linguistic concepts, and as such, the number of metaphors discovered is limited by the limited number of grounded concepts we do have. Also, dearth of a suitable dependency parser to automatically annotate relations between verbs and it's predicates creates noise in the system, which, in this work, has been avoided only through manual intervention. Provided this step is managed, the system can be fully automatized as a cognitive system capable of metaphor handling based on it's interaction with it's limited environment. Instead of 81 seconds of learning, however, the human learner has days and months and years of exposure, and clearly this can lead to the construction of extremely rich and diverse schemata. In the context of computational applications of language, such schema can be maintained much more easily than most traditional systems and provide a simple mechanism for updating world ontologies in an empirically validated manner. Through this paper, we have only hoped to present a new approach which might be truer to the original goal of AI for creating intelligent systems capable of human-like behavior.

References

Bailey, D. 1997. A computational model of embodiment in the acquisition of action verbs.

Ballard, D. H., and Yu, C. 2003. A multimodal learning interface for word acquisition. In *Proc. of ICASSP*.

Bloom, P. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Fass, D. 1991. Met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics* 17(1):49–90.

Feinerer, I., and Hornik, K. 2011. *wordnet: WordNet Interface*. R package version 0.1-8.

Heider, F., and Simmel, M. 1944. An experimental study of apparent behavior. *American Journal of Psychology* 57:243–259.

Kintsch, W. 2000. Metaphor comprehension: A computational theory. *Psychonomic Bulletin and Review* 257–266.

Lakoff, G., and Johnson, M., eds. 1980. *Metaphors We Live By*. University of Chicago Press.

Lakoff, G.; Espenson, J.; and Schwartz, A. 1991. *Master metaphor list: 2nd draft copy*.

Langacker, R., ed. 1987. *Foundations of Cognitive Grammar I: Theoretical Prerequisites*. Stanford University Press.

Mandler, J. M. 2004. *Foundations of Mind*. Oxford University Press.

Marie-Catherine de Marneffe, B. M., and Manning, C. D. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC 2006*.

Mason, Z. J. 2004. Cormet: A computational, corpus-based conventional metaphor extraction system. *Computational Linguistics* 30:23–44.

Mukerjee, A., and Sarkar, M. 2007. Grounded acquisition of containment prepositions. In *Proc. ICON*.

Mukerjee, A.; Neema, K.; and Nayak, S. 2011. Discovering coreference using image-grounded verb models. In *RANLP*, 610–615.

Narayanan, S. 1997. *Knowledge-based Action Representations for Metaphor and Aspect (KARMA)*. Ph.D. Dissertation, CS, UC Berkeley.

Narayanan, S. 1999. Moving right along: A computational model of metaphoric reasoning about events. In *Proc. of AAAI*, 121–129.

Nayak, S., and Mukerjee, A. 2012. Learning containment metaphors. In *Proc. of Annual Conference of the Cognitive Science Society*.

Resnik, P. S. 1993. Selection and information: A class-based approach to lexical relationships. Technical report.

Roy, D., and Reiter, E. 2005. Connecting language to the world. *Artificial Intelligence: Special Issue on Connecting Language to the World* 167:112.

Roy, D.; Hsiao, K.; and Mavridis, N. 2004. Mental imagery for a conversational robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 34(3):1374–1383.

Shutova, E., and Teufel., S. 2010. Metaphor corpus annotated for source - target domain mappings. In *Proceedings of LREC 2010*.

Shutova, E.; Sun, L.; and Korhonen, A. 2010. Metaphor identification using verb and noun clustering. In *Proc. of COLING*, 1002–1010.

Siskind, J. 1996. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61:39–91.

Solan, Z.; Ruppin, E.; Horn, D.; and Edelman, S. 2002. Automatic acquisition and efficient representation of syntactic structures. In *Proc. of NIPS*.

Wilks, Y. 1975. A preferential pattern seeking semantics for natural language inference. *Artificial Intelligence* 6(1):53–74.