# Minimally Supervised Multilingual Taxonomy and Translation Lexicon Induction

**Nikesh Garera** and **David Yarowsky**
Department of Computer Science
Center for Language and Speech Processing
Johns Hopkins University
Baltimore, MD 21218, USA
{ngarera,yarowsky}@cs.jhu.edu

## Abstract

We present a novel algorithm for the acquisition of multilingual lexical taxonomies (including hyponymy/hypernymy, meronymy and taxonomic cousinhood), from monolingual corpora with minimal supervision in the form of seed exemplars using discriminative learning across the major WordNet semantic relationships. This capability is also extended robustly and effectively to a second language (Hindi) via cross-language projection of the various seed exemplars. We also present a novel model of translation dictionary induction via multilingual transitive models of hypernymy and hyponymy, using these induced taxonomies. Candidate lexical translation probabilities are based on the probability that their induced hyponyms and/or hypernyms are translations of one another. We evaluate all of the above models on English and Hindi.

## 1 Introduction

Taxonomy resources such as WordNet are limited or non-existent for most of the world's languages. Building a WordNet manually from scratch requires a huge amount of human effort and for rare languages the required human and linguistic resources may simply not be available. Most of the automatic approaches for extracting semantic relations (such as hyponyms) have been demonstrated for English and some of them rely on various language-specific resources (such as supervised training data, language-specific lexicosyntactic patterns, shallow parsers, etc.). This paper presents a language independent approach for inducing taxonomies such as shown in Figure 1 using limited supervision and linguistic resources. We propose a seed learning based approach for extracting semantic relations (hyponyms, meronyms and cousins) that improves upon existing induction frameworks by combining evidence from multiple semantic relation types. We show that using a joint model for extracting different semantic relations helps to induce more relation-specific patterns and filter out the generic patterns[1]. The pat-

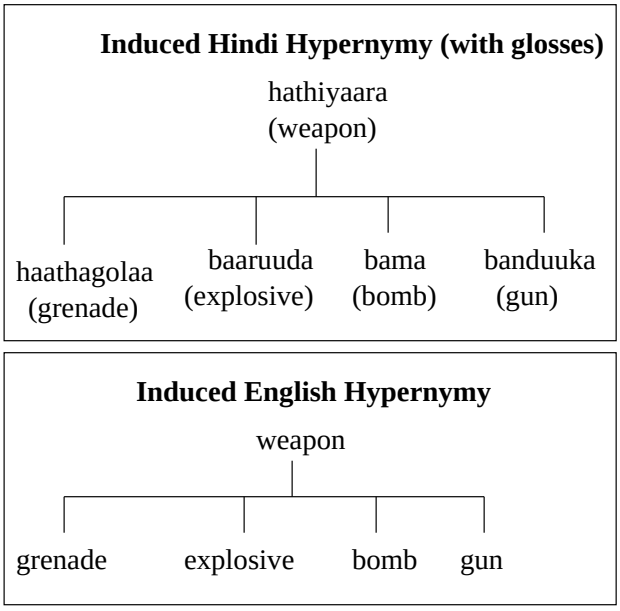


Figure 1: Goal: To induce multilingual taxonomy relationships in parallel in multiple languages (such as Hindi and English) for information extraction and machine translation purposes.

[1]By generic patterns, we mean patterns that cannot distinguish between different semantic relations. For example, the

terns can then be used for extracting new wordpairs expressing the relation. Note that the only training data used in the algorithm are the few seed pairs required to start the bootstrapping process, which are relatively easy to obtain. We evaluate the taxonomy induction algorithm on English and a second language (Hindi) and show that it can reliably and accurately induce taxonomies in two diverse languages.
We further show how having induced parallel taxonomies in two languages can be used for augmenting a translation dictionary between those two languages. We make use of the automatically induced hyponym/hypernym relations in each language to create a transitive "bridge" for dictionary induction. Specifically, the dictionary induction task relies on the key observation that words in two languages (e.g. English and Hindi) have increased probabilities of being translations of each other if their hypernyms or hyponyms are translations of one another.

## 2 Related Work

While manually created WordNets for English (Fellbaum, 1998) and Hindi (Narayan, 2002) have been made available, a lot of time and effort is required in building such semantic taxonomies from scratch. Hence several automatic corpus based approaches for acquiring lexical knowledge have been proposed in the literature. Much of this work has been done for English based on using a few evocative fixed patterns including "X and other Ys", "Y such as X", as in the classic work by Hearst (1992). The problems with using a few fixed patterns is the often low coverage of such patterns; thus there is a need for discovering additional informative patterns automatically. There has been a plethora of work in the area of information extraction using automatically derived patterns contextual patterns for semantic categories (e.g. companies, locations, time, person-names, etc.) based on bootstrapping from a small set of seed words (Riloff and Jones, 1999; Agichtein and Gravano, 2000; Thelen and Riloff, 2002; Ravichandran and Hovy, 2002; Hasegawa et al. 2004; Etzioni et al. 2005; Paşca et al. 2006). This framework has been also shown to work for extracting semantic relations between entities: Pantel et al. (2004) proposed an approach based on edit-distance to learn lexico-POS patterns for *is-a* and *part-of* relations. Girju et al. (2003) used 100 seed words from WordNet to extract patterns for *part-of* relations. While most of the above pattern induction work has been shown to work well for specific relations (such as "birthdates, companies, etc."), Section 3.1 explains why directly applying seed learning for semantic relations can result in high recall but low precision patterns, a problem also noted by Pantel and Pennacchiotti (2006). Furthermore, much of the semantic relation extraction work has focused on extracting a particular relation *independently* of other relations. We show how this problem can be solved by combining evidence from multiple relations in Section 3.2. Snow et al.(2006) also describe a probablistic framework for combining evidence using constraints from hyponymy and cousin relations. However, they use a supervised logistic regression model. Moreover, their features rely on parsing dependency trees which may not be available for most languages.
The key contribution of this work is using evidence from multiple relationship types in the seed learning framework for inducing these relationships and conducting a multilingual evaluation for the same. We further show how extraction of semantic relations in multiple languages can be applied to the task of improving a dictionary between those languages.

## 3 Approach

To be able to automatically create taxonomies such as WordNet, it is useful to be able to learn not only hyponymy/hyponymy directly, but also the additional semantic relationships of meronymy and taxonomic cousinhood. Specifically, given a pair of words (X, Y), the task is to answer the following questions: 1. Is X a hyponym of Y (e.g. *weapon, gun*)? 2. Is X a part/member of Y (e.g. *trigger, gun*)? 3. Is X a cousin/sibling[2] of Y (e.g. *gun, missile*)? 4. Do none of the above 3 relations apply but X is observed in the context of Y (e.g. *airplane,accident*)?[3] We will refer to class 4 as "other".

pattern "X and Y" is a generic pattern whereas the pattern "Y such as X" is a hyponym-specific pattern

[2]Cousins/siblings are words that share a close common hypernym

[3]Note that this does not imply X is unrelated or independent of Y. On the contrary, the required sentential co-occurence implies a topic similarity. Thus, this is a much harder class to distinguish from classes 1-3 than non co-occuring unrelatedness (such as *gun, protazoa*) and hence was included in the evaluation.

| Rank | English | Hindi |
|---|---|---|
| 1 | Y, the X | Y aura X (Gloss: Y and X) |
| 2 | Y and X | Y va X (Gloss: Y in addition to X) |
| 3 | X and other Y | Y ne X (Gloss: Y (case marker) X) |
| 4 | X and Y | X ke Y (Gloss: X's Y) |
| 5 | Y, X | Y me.n X (Gloss: Y in X) |

Table 1: Naive pattern scoring: Hyponymy patterns ranked by their raw corpus frequency scores.

| Rank | English | Hindi |
|---|---|---|
| 1 | Y like X | X aura anya Y (Gloss: X and other Y) |
| 2 | Y such as X | Y, X (Gloss: Y, X) |
| 3 | X and other Y | X jaise Y (Gloss: X like Y) |
| 4 | Y and X | Y tathaa X (Gloss: Y or X) |
| 5 | Y, including X | X va anya Y (Gloss: X and other Y) |

Table 2: Patterns for hypernymy class reranked using evidence from other classes. Patterns distributed fairly evenly across multiple relationship types (e.g. "X and Y") are deprecated more than patterns focused predominantly on a single relationship type (e.g. "Y such as X").

### 3.1 Independently Bootstrapping Lexical Relationship Models

Following the pattern induction framework of Ravichandran and Hovy (2002), one of the ways of extracting different semantic relations is to learn patterns for each relation independently using seeds of that relation and extract new pairs using the learned patterns. For example, to build an independent model of hyponymy using this framework, we collected approximately 50 seed exemplars of hyponym pairs and extracted all the patterns that match with the seed pairs[4]. As in Ravichandran and Hovy (2002), the patterns were ranked by corpus frequency and a frequency threshold was set to select the final patterns. These patterns were then used to extract new word pairs expressing the hyponymy relation by finding word pairs that occur with these patterns in an unlabeled corpus. However, the problem with this approach is that generic patterns (like "X and Y") occur many times in a corpus and thus low-precision patterns may end up with high cumulative scores. This problem is illustrated more clearly in Table 1, which shows a list of top five hyponymy patterns (ranked by their corpus frequency) using this approach. We overcome this problem by exploiting the multi-class nature of our task and combine evidence from multiple relations in order to learn high precision patterns (with high conditional probabilities) for each relation. The key idea is to weed out the patterns that occur in more than one semantic relation and keep the ones that are relation-specific[5], thus using the relations meronymy, cousins and other as *negative evidence* for hyponymy and vice versa. Table 2 shows the pattern ranking by using the model developed in Section 3.2 that makes use of evidence from different classes. We can see more hyponymy specific patterns ranked at the top[6] suggesting the usefulness of this method in finding class-specific patterns.

### 3.2 A minimally supervised multi-class classifier for identifying different semantic relations

First, we extract a list of patterns from an unlabeled corpus[7] independently for each relationship type (class) using the seeds[8] for the respective class as in Section 3.1.[9] In order to develop a multi-

[4] A pattern is the ngrams occurring between the seedpair (also called gluetext). The length of the pattern was thresholded to 15 words.

[5] In the actual algorithm, we will not be entirely weeding out the common patterns but will estimate the conditional class probabilities for each pattern: $p(class|pattern)$

[6] It is interesting to see in Table 2 that the top learned Hindi hyponymy patterns seem to be translations of the English patterns suggested by Hearst (1992). This leads to an interesting future work question: Are the most effective hyponym patterns in other languages usually translations of the English hyponym patterns proposed by Hearst (1992) and what are frequent exceptions?

[7] Unlabeled monolingual corpora were used for this task, the English corpus was the LDC Gigaword corpus and the Hindi corpus was newswire text extracted from the web containing a total of 64 million words.

[8] The number of seeds used for classes {hyponym, meronym, cousin, other} were {48,40,49,50} for English and were {32,58,31,35} for Hindi respectively. A sample of seeds used is shown in Table 5.

[9] We retained only the patterns that had seed frequency greater one for extracting new word pairs. The total number

| | Hypo. | Mero. | Cous. | Other |
|---|---|---|---|---|
| X *of the Y* | 0 | 0.66 | 0.04 | 0.3 |
| Y, *especially X* | 1 | 0 | 0 | 0 |
| Y, *whose X* | 0 | 1 | 0 | 0 |
| X *and other Y* | 0.63 | 0.08 | 0.18 | 0.11 |
| X *and Y* | 0.23 | 0.3 | 0.33 | 0.14 |

Table 3: A sample of patterns and their relationship type probabilities $P(class|pattern)$ extracted at the end of training phase for English.

| | Hypo. | Mero. | Cous. | Other |
|---|---|---|---|---|
| X *aura anya Y* (X and other Y) | 1 | 0 | 0 | 0 |
| X *aura Y* (X and Y) | 0.09 | 0.09 | 0.71 | 0.11 |
| X *jaise Y* (X like Y) | 1 | 0 | 0 | 0 |
| X *va Y* (X and Y) | 0.11 | 0 | 0.89 | 0 |
| Y *kii X* (Y's X) | 0.33 | 0.67 | 0 | 0 |

Table 4: A sample of patterns and their class probabilities $P(class|pattern)$ extracted at the end of training phase for Hindi.

class probabilistic model, we obtain the probability of each class $c$ given the pattern $p$ as follows:

$$P(c|p) = \frac{seed_{freq}(p,c)}{\sum_{c'} seed_{freq}(p,c')}$$

where $seed_{freq}(p, c)$ is the number of seeds of class c that were found with the pattern p in an unlabeled corpus. A sample of the $P(class|pattern)$ tables for English and Hindi are shown in the Tables 3 and 4 respectively. It is clear how occurrence of a pattern in multiple classes can be used for finding reliable patterns for a particular class. For example, in Table 3: although the pattern "X and Y" will get a higher seed frequency than the pattern "Y, especially X", the probability $P(\text{“}X\ and\ Y''|hyponymy)$ is much lower than $P(\text{“}Y,\ especially\ X''|hyponymy)$, since the pattern "Y, especially X" is unlikely to occur with seeds of other relations.

Now, instead of using the $seed_{freq}(p, c)$ as the score for a particular pattern with respect to a class, we can rescore patterns using the probabilities $P(class|pattern)$. Thus the final score for a pattern p with respect to class c is obtained as:

$$score(p, c) = seed_{freq}(p, c) \cdot P(c|p)$$

We can view this equation as balancing recall and precision, where the first term is the frequency of the pattern with respect to seeds of class c (representing recall), and the second term represents the relation-specificness of the pattern with respect to class c (representing precision). We recomputed the score for each pattern in the above manner and obtain a ranked list of patterns for each of the classes for English and Hindi. Now, to extract new pairs for each class, we take all the patterns with a seed frequency greater than 2 and use them to extract word pairs from an unlabeled corpus. The semantic class for each extracted pair is then predicted using the multi-class classifier as follows: Given a pair of words (X1, X2), note all the patterns that matched with this pair in the unlabeled corpus, denote this set as $\mathcal{P}$. Choose the predicted class $c*$ for this pair as:

$$c^* = argmax_c \sum_{p \in \mathcal{P}} score(p, c)$$

### 3.3 Evaluation of the Classification Task

Over 10,000 new word relationship pairs were extracted based on the above algorithm. While it is hard to evaluate all the extracted pairs manually, one can certainly create a representative smaller test set and evaluate performance on that set. The test set was created by randomly identifying word pairs in WordNet and newswire corpora and annotating their correct semantic class relationships. Test set construction was done entirely independently from the algorithm application, and hence some of the test pairs were missed entirely by the learning algorithm, yielding only partial coverage.

The total number of test examples including all classes were 200 and 140 for English and Hindi test-sets respectively. The overall coverage[10] on these test-sets was 81% and 79% for English and Hindi respectively. Table 6 reports the overall accuracy[11] for the 4-way classification using different patterns scoring methods. Baseline 1 is scoring patterns by their corpus frequency as in Ravichandran and Hovy (2002), Baseline 2 is another intutive method of

of retained patterns across all classes for {English,Hindi} were {455,117} respectively.

[10] Coverage is defined as the percentage of the test cases that were present in the unlabeled corpus, that is, cases for which an answer was given.

[11] Accuracy on a particular set of pairs is defined as the percentage of pairs in that set whose class was correctly predicted.

| | English | | Hindi | |
|---|---|---|---|---|
| | Seed Pairs | Model Predictions | Seed Pairs | Model Predictions |
| Hypernym | tool,hammer | gun,weapon | khela,Tenisa (game,tennis) | kaa.ngresa,paarTii (congress,party) |
| | currency,yen | hockey,sport | appraadha,hatyaa (crime,murder) | passporTa,kaagajaata (passport,document) |
| | metal,copper | cancer,disease | jaanvara,bhaaga (animal,tiger) | a.ngrejii,bhaashhaa (English,language) |
| Meronym | wheel,truck | room,hotel | u.ngalii,haatha (finger,hand) | jeba,sharTa (pocket,shirt) |
| | headline,newspaper | bark,tree | kamaraa,aspataala (room,hospital) | kaptaana,Tiima (captain,team) |
| | wing,bird | lens,camera | ma.njila,imaarata (floor,building) | darvaaja,makaana (door,house) |
| Cousin | dollar,euro | guitar,drum | bhaajapa,kaa.ngresa (bjp,congress) | peTrola,Diijala (petrol,diesel) |
| | heroin,cocaine | history, geography | Hindii,a.ngrejii (Hindi,English) | Daalara,rupayaa (dollar,rupee) |
| | helicopter,submarine | diabetes,arthritis | basa,Traka (bus,truck) | talaaba,nadii (pond,river) |

Table 5: A sample of seeds used and model predictions for each class for the taxonomy induction task. For each of the model predictions shown above, its Hyponym/Meronym/Cousin classification was correctly assigned by the model.

scoring patterns by the number of seeds they extract. The third row in Table 6 indicates the result of rescoring patterns by their class conditional probabilties, giving the best accuracy.

While this method yields some improvement over other baselines, the main point to note here is that the pattern-based methods which have been shown to work well for English also perform reasonably well on Hindi, inspite of the fact that the size of the unlabeled corpus available for Hindi was 15 times smaller than for English.

Table 7 shows detailed accuracy results for each relationship type using the model developed in section 3.2. It is also interesting to see in Table 8 that most of the confusion is due to "other" class being classified as "cousin" which is expected as cousin words are only weakly semantically related and uses more generic patterns such as "X and Y" which can often be associated with the "other" class as well. Strongly semantically clear classes like Hypernymy and Meronymy seem to be well discriminated as their induced patterns are less likely to occur in other relationship types.

| Model | English Accuracy | Hindi Accuracy |
|---|---|---|
| Baseline 1 $_{[RH02]}$ | 65% | 63% |
| Baseline 2 $_{seed_{freq}}$ | 70% | 65% |
| $seed_{freq} \cdot P(c\|p)$ | **73%** | **66%** |

Table 6: Overall accuracy for 4-way classification {hypernym,meronym,cousin,other} using different pattern scoring methods.

| | English | | | Hindi | | |
|---|---|---|---|---|---|---|
| | Total | Cover. | Acc. | Total | Cover. | Acc. |
| Hypr. | 83 | 74% | 97% | 59 | 82% | 75% |
| Mero. | 41 | 81% | 88% | 33 | 63% | 81% |
| Cous. | 42 | 91% | 55% | 23 | 91% | 71% |
| Other | 34 | 85% | 31% | 25 | 80% | 20% |
| Overall | 200 | 81% | 73% | 140 | 79% | 66% |

Table 7: Test set coverage and accuracy results for inducing different semantic relationship types.

| | English | | | | Hindi | | | |
|---|---|---|---|---|---|---|---|---|
| | Hypo. | Mero. | Cous. | Oth. | Hypo. | Mero. | Cous. | Oth. |
| Hypo. | 59 | 1 | 1 | 0 | 36 | 1 | 10 | 1 |
| Mero. | 1 | 28 | 1 | 3 | 0 | 17 | 4 | 0 |
| Cous. | 14 | 3 | 21 | 0 | 6 | 0 | 15 | 0 |
| Other | 7 | 3 | 10 | 9 | 1 | 4 | 11 | 4 |

Table 8: Confusion matrix for English (left) Hindi (right) for the four-way classification task

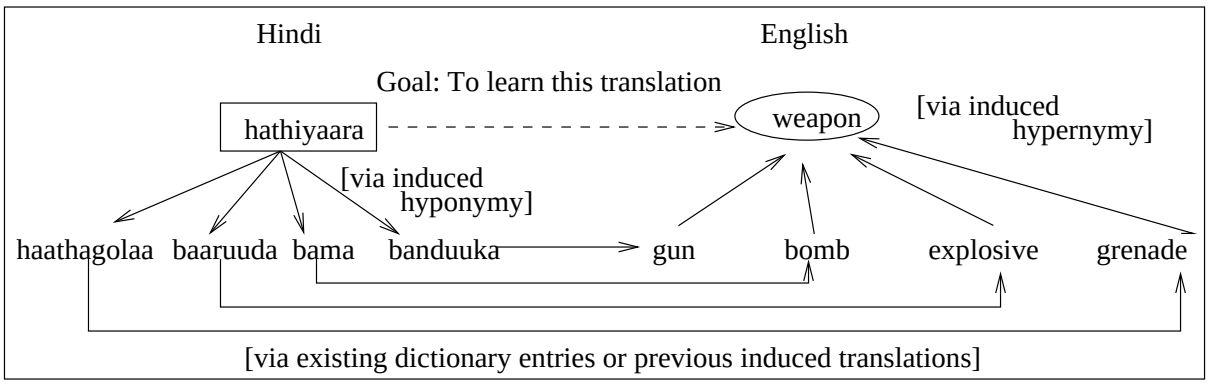


Figure 2: Illustration of the models of using induced hyponymy and hypernymy for translation lexicon induction.

## 4 Improving a partial translation dictionary

In this section, we explore the application of automatically generated multilingual taxonomies to the task of translation dictionary induction. The hypothesis is that a pair of words in two languages would have increased probability of being translations of each other if their hypernyms or hyponyms are translations of one another.

As illustrated in Figure 2, the probability that *weapon* is a translation of the Hindi word *hathiyaara* can be decomposed into the sum of the probabilities that their hyponyms in both languages (as induced in Section 3.2) are translations of each other. Thus:

$$P_{H->E}\ (W_E|W_H) = \sum_i P_{hyper}\ (W_E|Eng(H_i))\ P_{hypo}(H_i|W_H)$$

for induced hyponyms $H_i$ of the source word $W_H$, and using an existing (and likely very incomplete) Hindi-English dictionary to generate Eng($H_i$) for these hyponyms, and the corresponding induced hypernyms of these translations in English.[12]. We conducted a very preliminary evaluation of this idea for obtaining English translations of a set of 25 Hindi words. The Hindi candidate hyponym space had been pruned of function words and non-noun words. The likely English translation candidates for each Hindi word were ranked according to the probability $P_{H->E}(W_E|W_H)$.

The first column of Table 9 shows the stand-alone performance for this model on the dictionary induction task. This standalone model has a reasonably good accuracy for finding the correct translation in the Top 10 and Top 20 English candidates.

| | Accuracy (uni-d) | Accuracy (bi-d) | Accuracy bi-d + Other |
|---|---|---|---|
| Top 1 | 20% | 36% | 36% |
| Top 5 | 56% | 64% | 72% |
| Top 10 | 72% | 72% | 80% |
| Top 20 | 84% | 84% | 84% |

Table 9: Accuracy on Hindi to English word translation using different transitive hypernym algorithms. The additional model components in the bi-d(irectional) plus Other model are only used to rerank the top 20 candidates of the bidirectional model, and are hence limited to its top-20 performance.

This approach can be further improved by also implementing the above model in the reverse direction and computing the $P(W_H|W_{E_i})$ for each of the English candidates $E_i$. We did so and computed $P(W_H|W_{E_i})$ for top 20 English candidate translations. The final score for an English candidate translation given a Hindi word was combined by a simple average of the two directions, that is, by summing $P(W_{E_i}|W_H)$ + $P(W_H|W_{E_i})$.

The second column of Table 9 shows how this bidirectional approach helps in getting the right

[12]One of the challenges of inducing a dictionary via using a corpus based taxonomy is sense disambiguation of the words to be translated. In the current model, the more dominant sense (in terms of corpus frequency of its hyponyms) is likely to get selected by this approach. While the current model can still help in getting translations of the dominant sense, possible future work would be to cluster all the hyponyms according to contextual features such that each cluster can represent the hyponyms for a particular sense. The current dictionary induction model can then be applied again using the hyponym clusters to distinguish different senses for translation.

translations in Top 1 and Top 5 as compared to the unidirectional approach. Table 10 shows a sample

| Correctly translated | Incorrectly translated |
|---|---|
| aujaara (tool) | vishaya (topic) |
| biimaarii (disease) | saamana (stuff) |
| hathiyaara (weapon) | dala (group,union) |
| dastaaveja (documents) | tyohaara (festival) |
| aparaadha (crime) | jagaha (position,location) |

Table 10: A sample of correct and incorrect translations using transitive hypernymy/hyponym word translation induction

of correct and incorrect translations generated by the above model. It is interesting to see that the incorrect translations seem to be the words that are very general (like "topic", "stuff", etc.) and hence their hyponym space is very large and diffuse, resulting in incorrect translations.While the columns 1 and 2 of Table 9 show the standalone application of our translation dictionary induction method, we can also combine our model with existing work on dictionary induction using other translation induction measures such as using relative frequency similarity in multilingual corpora and using cross-language context similarity between word co-occurrence vectors (Schafer and Yarowsky, 2002).We implemented the above dictionary induction measures and combined the taxonomy based dictionary induction model with other measures by just summing the two scores[13]. The preliminary results for bidirectional hypernym/hyponym + other features are shown in column 3 of Table 9. The results show that the hypernym/hyponym features can be a useful orthogonal source of lexical similarity in the translation-induction model space. While the model shown in Figure 2 proposes inducing translations of hypernyms, one can also go in the other direction and induce likely translation candidates for hyponyms by knowing the translation of hypernyms. For example, to learn that *rifle* is a likely translation candidate of the Hindi word *raaiphala*, is illustrated in Figure 3. But because there is a much larger space of hyponyms for *weapon* in this direction, the output serves more to reduce the entropy of the translation candidate space when used in conjunction with other translation induction similarity measures. We would expect the application of additional similarity measures to this greatly narrowed and ranked hypothesis space to yield improvement in future work.

[13]after renormalizing each of the individual score to be in the range 0 to 1.

## 5 Conclusion

This paper has presented a novel minimal-resource algorithm for the acquisition of multilingual lexical taxonomies (including hyponymy/hypernymy and meronymy). The algorithm is based on cross language projection of various monolingual indicators of these taxonomic relationships in free text and via bootstrapping thereof. Using only 31-58 seed examples, the algorithm achieves accuracies of 73% and 66% for English and Hindi respectively on the tasks of hyponymy/meronomy/cousinhood/other model induction. The robustness of this approach is shown by the fact that the unannotated Hindi development corpus was only 1/15th the size of the utilized English corpus. We also present a novel model of unsupervised translation dictionary induction via multilingual transitive models of hypernymy and hyponymy, using these induced taxonomies and evaluated on Hindi-English. Performance starting from no multilingual dictionary supervision is quite promising.

## References

E. Agichtein and L. Gravano. 2000. Snowball: extracting relations from large plain-text collections. In *Proceedings of the 5th ACM International Conference on Digital Libraries*, pages 85–94.

M. J. Cafarella, D. Downey, S. Soderland, and O. Etzioni. 2005. Knowitnow: Fast, scalable information extraction from the web. In *Proceedings of EMNLP/HLT-05*, pages 563–570.

S. Caraballo. 1999. Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proceedings of ACL-99*, pages 120–126.

B. Carterette, R. Jones, W. Greiner, and C. Barr. 2006. N semantic classes are harder than two. In *Proceedings of ACL/COLING-06*, pages 49–56.

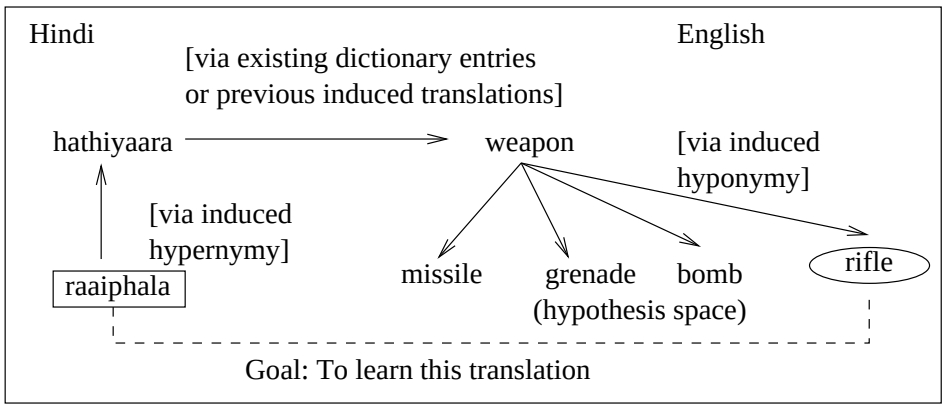


Figure 3: Reducing the space of likely translation candidates of the word *raaiphala* by inducing its hypernym, using a partial dictionary to look up the translation of hypernym and generating the candidate translations as induced hyponyms in English space.


O. Etzioni, M. Cafarella, D. Downey, A. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates. 2005. Unsupervised named-entity extraction from the web: an experimental study. *Artif. Intell.*, 165(1):91–134.

C. Fellbaum. 1998. *WordNet: An electronic lexical database.*

*R. Girju, A. Badulescu, and D. Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In* Proceedings of HLT/NAACL-03, *pages 1–8.*

*R. Girju, A. Badulescu, and D. Moldovan. 2006. Automatic discovery of part-whole relations.* Computational Linguistics, *21(1):83–135.*

*D. Graff, J. Kong, K. Chen, and K. Maeda. 2005. English Gigaword Second Edition. Linguistic Data Consortium, catalog number LDC2005T12.*

*T. Hasegawa, S. Sekine, and R. Grishman. 2004. Discovering relations among named entities from large corpora. In* Proceedings of ACL-04, *pages 415–422.*

*M. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In* Proceedings of COLING-92, *pages 539–545.*

*D. Narayan, D. Chakrabarty, P. Pande, and P. Bhattacharyya. 2002. An Experience in Building the Indo WordNet-a WordNet for Hindi.* International Conference on Global WordNet.

*Marius Paşca, Dekang Lin, Jeffrey Bigham, Andrei Lifchits, and Alpa Jain. 2006. Names and similarities on the web: Fact extraction in the fast lane. In* Proceedings of ACL/COLING-06, *pages 809–816.*

*P. Pantel and M. Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In* Proceedings of ACL/COLING-06, *pages 113–120.*

*P. Pantel and D. Ravichandran. 2004. Automatically labeling semantic classes. In* Proceedings of HLT/NAACL-04, *pages 321–328.*

*P. Pantel, D. Ravichandran, and E. Hovy. 2004. Towards terascale knowledge acquisition. In* Proceedings of COLING-04.

*D. Ravichandran and E. Hovy. 2002. Learning surface text patterns for a question answering system. In* Proceedings of ACL-02, *pages 41–47.*

*E. Riloff and R. Jones. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In* Proceedings of AAAI/IAAI-99, *pages 474–479.*

*E. Riloff and J. Shepherd. 1997. A corpus-based approach for building semantic lexicons.* CoRR, *cmp-lg/9706013.*

*C. Schafer and D. Yarowsky. 2002. Inducing translation lexicons via diverse similarity measures and bridge languages. In* Proceedings of CONLL-02, *pages 146–152.*

*R. Snow, D. Jurafsky, and A. Y. Ng. 2006. Semantic taxonomy induction from heterogenous evidence. In* Proceedings of ACL/COLING-06, *pages 801–808.*

*M. Thelen and E. Riloff. 2002. A bootstrapping method for learning semantic lexicons using extraction pattern contexts. In* Proceedings of EMNLP-02, *pages 214–221.*

*D. Widdows. 2003. Unsupervised methods for developing taxonomies by combining syntactic and statistical information. In* Proceedings of HLT/NAACL-03, *pages 197–204.*