Don't 'have a clue'?
Unsupervised co-learning of downward-entailing operators

pp. 1-6 are identical to the ACL 2010 published version; pp. 7-8 are the "externally-available appendices".

Cristian Danescu-Niculescu-Mizil and Lillian Lee
Department of Computer Science, Cornell University
cristian@cs.cornell.edu, llee@cs.cornell.edu

Abstract

Researchers in textual entailment have begun to consider inferences involving *downward-entailing operators*, an interesting and important class of lexical items that change the way inferences are made. Recent work proposed a method for learning English downward-entailing operators that requires access to a high-quality collection of *negative polarity items* (NPIs). However, English is one of the very few languages for which such a list exists. We propose the first approach that can be applied to the many languages for which there is no pre-existing high-precision database of NPIs. As a case study, we apply our method to Romanian and show that our method yields good results. Also, we perform a cross-linguistic analysis that suggests interesting connections to some findings in linguistic typology.

1 Introduction

Cristi: *"Nicio" ... is that adjective you've mentioned.*
Anca: *A negative pronominal adjective.*
Cristi: *You mean there are people who analyze that kind of thing?*
Anca: *The Romanian Academy.*
Cristi: *They're crazy.*
—From the movie *Police, adjective*

Downward-entailing operators are an interesting and varied class of lexical items that change the default way of dealing with certain types of inferences. They thus play an important role in understanding natural language [6, 18–20, etc.].

We explain what downward entailing means by first demonstrating the "default" behavior, which is *upward entailing*. The word 'observed' is an example upward-entailing operator: the statement

(i) 'Witnesses <u>observed</u> opium use.'

implies

(ii) 'Witnesses <u>observed</u> narcotic use.'

but not vice versa (we write i $\Rightarrow$ ($\not\Leftarrow$) ii). That is, the truth value is preserved if we replace the argument of an upward-entailing operator by a superset (a more general version); in our case, the set 'opium use' was replaced by the superset 'narcotic use'.

Downward-entailing (DE) (also known as *downward monotonic* or *monotone decreasing*) operators violate this default inference rule: with DE operators, reasoning instead goes from "sets to subsets". An example is the word 'bans':

'The law $\boxed{\text{bans}}$ opium use'
$\not\Rightarrow$ ($\Leftarrow$)
'The law $\boxed{\text{bans}}$ narcotic use'.

Although DE behavior represents an exception to the default, DE operators are as a class rather common. They are also quite diverse in sense and even part of speech. Some are simple negations, such as 'not', but some other English DE operators are 'without', 'reluctant to', 'to doubt', and 'to allow'.[1] This variety makes them hard to extract automatically.

Because DE operators violate the default "sets to supersets" inference, identifying them can potentially improve performance in many NLP tasks. Perhaps the most obvious such tasks are those involving textual entailment, such as question answering, information extraction, summarization, and the evaluation of machine translation [4]. Researchers are in fact beginning to build textual-entailment systems that can handle inferences involving downward-entailing operators other than simple negations, although these systems almost all rely on small handcrafted lists of DE operators [1–3, 15, 16].[2] Other application areas are natural-language generation and human-computer interaction, since downward-entailing inferences induce

[1] Some examples showing different constructions for analyzing these operators: 'The defendant does not own a blue car' $\not\Rightarrow$ ($\Leftarrow$) 'The defendant does not own a car'; 'They are reluctant to tango' $\not\Rightarrow$ ($\Leftarrow$) 'They are reluctant to dance'; 'Police doubt Smith threatened Jones' $\not\Rightarrow$ ($\Leftarrow$) 'Police doubt Smith threatened Jones or Brown'; 'You are allowed to use Mastercard' $\not\Rightarrow$ ($\Leftarrow$) 'You are allowed to use any credit card'.

[2] The exception [2] employs the list automatically derived by Danescu-Niculescu-Mizil, Lee, and Ducott [5], described later.

greater cognitive load than inferences in the opposite direction [8].

Most NLP systems for the applications mentioned above have only been deployed for a small subset of languages. One key factor is the lack of relevant resources for other languages. While one approach would be to separately develop a method to acquire such resources for each language individually, we instead aim to ameliorate the resource-scarcity problem wholesale in the case of DE operators: we propose a single unsupervised method that can extract DE operators in any language for which raw text corpora exist.

Overview of our work Our approach takes the English-centric work of Danescu-Niculescu-Mizil et al. [5] — DLD09 for short — as a starting point, as they present the first and, until now, only algorithm for automatically extracting DE operators from data. However, our work departs significantly from DLD09 in the following key respect.

DLD09 critically depends on access to a high-quality, carefully curated collection of *negative polarity items (NPIs)* — lexical items such as 'any', 'ever', or the idiom 'have a clue' that tend to occur only in negative environments (see §2 for more details). DLD09 use NPIs as signals of the occurrence of downward-entailing operators. However, almost every language other than English lacks a high-quality accessible NPI list.

To circumvent this problem, we introduce a knowledge-lean *co-learning* approach. Our algorithm is initialized with a very small seed set of NPIs (which we describe how to produce), and then iterates between (a) discovering a set of DE operators using a collection of *pseudo-NPIs* — a concept we introduce — and (b) using the newly-acquired DE operators to detect new pseudo-NPIs.

Why this isn't obvious Although the algorithmic idea sketched above seems quite simple, it is important to note that prior experiments in that direction have not proved fruitful. Preliminary work on learning (German) NPIs using a small list of simple known DE operators did not yield strong results [14]. Hoeksema [10] discusses why NPIs might be hard to learn from data.[3] We circumvent this problem because we are not interested in learning NPIs per se; rather, for our purposes, pseudo-NPIs suffice. Also, our preliminary work determined that one of the most famous co-learning algorithms, *hubs and authorities* or *HITS* [11], is poorly suited to our problem.[4]

Contributions To begin with, we apply our algorithm to produce the first large list of DE operators for a language other than English. In our case study on Romanian (§4), we achieve quite high precisions at k (for example, iteration achieves a precision at 30 of 87%).

Auxiliary experiments explore the effects of using a large but noisy NPI list, should one be available for the language in question. Intriguingly, we find that co-learning new pseudo-NPIs provides better results.

Finally (§5), we engage in some cross-linguistic analysis based on the results of applying our algorithm to English. We find that there are some suggestive connections with findings in linguistic typology.

Appendix available A more complete account of our work and its implications can be found in a version of this paper containing appendices, available at www.cs.cornell.edu/~cristian/acl2010/.

2 DLD09: successes and challenges

In this section, we briefly summarize those aspects of the DLD09 method that are important to understanding how our new co-learning method works.

DE operators and NPIs Acquiring DE operators is challenging because of the complete lack of annotated data. DLD09's insight was to make use of *negative polarity items (NPIs)*, which are words or phrases that tend to occur only in negative contexts. The reason they did so is that Ladusaw's hypothesis [7, 13] asserts that *NPIs only occur within the scope of DE operators*. Figure 1 depicts examples involving the English NPIs 'any'[5] and 'have a clue' (in the idiomatic sense) that illustrate this relationship. Some other English NPIs are 'ever', 'yet' and 'give a damn'.

Thus, NPIs can be treated as clues that a DE operator might be present (although DE operators may also occur without NPIs).

[3] In fact, even humans can have trouble identifying NPIS; for instance, Lichte and Soehn [14] mention doubts about over half of Kürschner [12]'s 344 manually collected German NPIs.

[4] We explored three different edge-weighting schemes based on co-occurrence frequencies and seed-set membership, but the results were extremely poor; HITS invariably retrieved very frequent words.

[5] The *free-choice* sense of 'any', as in 'I can skim any paper in five minutes', is a known exception.

DE operators	NPIs: any^3	NPIs: *have a clue*, idiomatic sense
$\boxed{\text{not}}$ or $\boxed{\text{n't}}$	✓ We do $\boxed{\text{n't}}$ have *any* apples	✓ We do $\boxed{\text{n't}}$ *have a clue*
$\boxed{\text{doubt}}$	✓I $\boxed{\text{doubt}}$ they have *any* apples	✓ I $\boxed{\text{doubt}}$ they *have a clue*
no DE operator	× They have *any* apples	× They *have a clue*

Figure 1: Examples consistent with Ladusaw's hypothesis that NPIs can only occur within the scope of DE operators. A ✓ denotes an acceptable sentence; a × denotes an unacceptable sentence.

DLD09 algorithm Potential DE operators are collected by extracting those words that appear in an NPI's context at least once.[6] Then, the potential DE operators x are ranked by

$$f(x) := \frac{\text{fraction of NPI contexts that contain } x}{\text{relative frequency of } x \text{ in the corpus}},$$

which compares x's probability of occurrence conditioned on the appearance of an NPI with its probability of occurrence overall.[7]

The method just outlined requires access to a list of NPIs. DLD09's system used a subset of John Lawler's carefully curated and "moderately complete" list of English NPIs.[8] The resultant rankings of candidate English DE operators were judged to be of high quality.

The challenge in porting to other languages: cluelessness Can the unsupervised approach of DLD09 be successfully applied to languages other than English? Unfortunately, for most other languages, it does not seem that large, high-quality NPI lists are available.

One might wonder whether one can circumvent the NPI-acquisition problem by simply translating a known English NPI list into the target language. However, NPI-hood need not be preserved under translation [17]. Thus, for most languages, we lack the critical clues that DLD09 depends on.

3 Getting a clue

In this section, we develop an iterative co-learning algorithm that can extract DE operators in the many languages where a high-quality NPI database is not available, using Romanian as a case study.

3.1 Data and evaluation paradigm

We used Rada Mihalcea's corpus of ≈1.45 million sentences of raw Romanian newswire articles.

Note that we cannot evaluate impact on textual inference because, to our knowledge, no publicly available textual-entailment system or evaluation data for Romanian exists. We therefore examine the system outputs directly to determine whether the top-ranked items are actually DE operators or not. Our evaluation metric is precision at k of a given system's ranked list of candidate DE operators; it is not possible to evaluate recall since no list of Romanian DE operators exists (a problem that is precisely the motivation for this paper).

To evaluate the results, two native Romanian speakers labeled the system outputs as being "DE", "not DE" or "Hard (to decide)". The labeling protocol, which was somewhat complex to prevent bias, is described in the externally-available appendices (§7.1). The complete system output and annotations are publicly available at: http://www.cs.cornell.edu/~cristian/acl2010/.

3.2 Generating a seed set

Even though, as discussed above, the translation of an NPI need not be an NPI, a preliminary review of the literature indicates that in many languages, there is some NPI that can be translated as 'any' or related forms like 'anybody'. Thus, with a small amount of effort, one can form a minimal NPI seed set for the DLD09 method by using an appropriate target-language translation of 'any'. For Romanian, we used 'vreo' and 'vreun', which are the feminine and masculine translations of English 'any'.

3.3 DLD09 using the Romanian seed set

We first check whether DLD09 with the two-item seed set described in §3.2 performs well on Romanian. In fact, the results are fairly poor:

[6] DLD09 policies: (a) "NPI context" was defined as the part of the sentence to the left of the NPI up to the first comma, semi-colon or beginning of sentence; (b) to encourage the discovery of new DE operators, those sentences containing one of a list of 10 well-known DE operators were discarded. For Romanian, we treated only negations ('nu' and 'n-') and questions as well-known environments.

[7] DLD09 used an additional *distilled* score, but we found that the distilled score performed worse on Romanian.

[8] http://www-personal.umich.edu/~jlawler/aue/npi.html

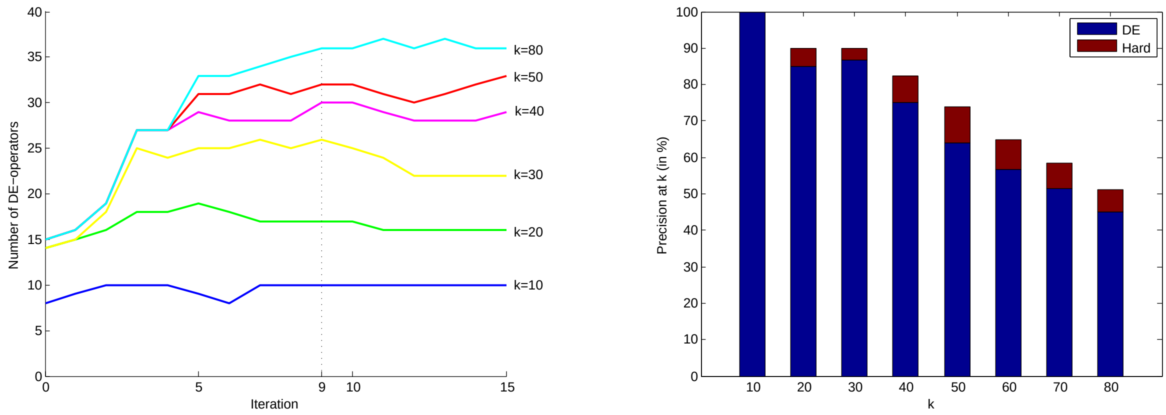

Figure 2: **Left:** Number of DE operators in the top k results returned by the co-learning method at each iteration. Items labeled "Hard" are not included. Iteration 0 corresponds to DLD09 applied to $\{$'vreo', 'vreun'$\}$. Curves for $k = 60$ and 70 omitted for clarity. **Right:** Precisions at k for the results of the 9th iteration. The bar divisions are: DE (blue/darkest/largest) and Hard (red/lighter, sometimes non-existent).

for example, the precision at 30 is below 50%. (See blue/dark bars in figure 3 in the externally-available appendices for detailed results.)

This relatively unsatisfactory performance may be a consequence of the very small size of the NPI list employed, and may therefore indicate that it would be fruitful to investigate automatically extending our list of clues.

3.4 Main idea: a co-learning approach

Our main insight is that not only can NPIs be used as clues for finding DE operators, as shown by DLD09, but conversely, DE operators (if known) can potentially be used to discover new NPI-like clues, which we refer to as *pseudo-NPIs* (or *pNPIs* for short). By "NPI-like" we mean, "serve as possible indicators of the presence of DE operators, regardless of whether they are actually restricted to negative contexts, as true NPIs are". For example, in English newswire, the words 'allegation' or 'rumor' tend to occur mainly in DE contexts, like 'denied' or 'dismissed', even though they are clearly not true NPIs (the sentence 'I heard a rumor' is fine). Given this insight, we approach the problem using an *iterative co-learning* paradigm that integrates the search for new DE operators with a search for new pNPIs.

First, we describe an algorithm that is the "reverse" of DLD09 (henceforth *rDLD*), in that it retrieves and ranks pNPIs assuming a given list of DE operators. Potential pNPIs are collected by extracting those words that appear in a DE context (defined here, to avoid the problems of parsing or scope determination, as the part of the sentence to the right of a DE operator, up to the first comma, semi-colon or end of sentence); these candidates x are then ranked by

$$f_r(x) := \frac{\text{fraction of DE contexts that contain } x}{\text{relative frequency of } x \text{ in the corpus}}.$$

Then, our *co-learning* algorithm consists of the iteration of the following two steps:

- (*DE learning*) Apply DLD09 using a set $\mathcal{N}$ of pseudo-NPIs to retrieve a list of candidate DE operators ranked by f (defined in Section 2). Let $\mathcal{D}$ be the top n candidates in this list.

- (*pNPI learning*) Apply rDLD using the set $\mathcal{D}$ to retrieve a list of pNPIs ranked by f_r; extend $\mathcal{N}$ with the top n_r pNPIs in this list. Increment n.

Here, $\mathcal{N}$ is initialized with the NPI seed set. At each iteration, we consider the output of the algorithm to be the ranked list of DE operators retrieved in the DE-learning step. In our experiments, we initialized n to 10 and set n_r to 1.

4 Romanian results

Our results show that there is indeed favorable synergy between DE-operator and pNPI retrieval. Figure 2 plots the number of correctly retrieved DE operators in the top k outputs at each iteration. The point at iteration 0 corresponds to a datapoint already discussed above, namely, DLD09 applied to the two 'any'-translation NPIs. Clearly, we see general substantial improvement over DLD09, although the increases level off in later iterations.

(Determining how to choose the optimal number of iterations is a subject for future research.)

Additional experiments, described in the externally-available appendices (§7.2), suggest that pNPIs can even be more effective clues than a noisy list of NPIs. (Thus, a larger seed set does not necessarily mean better performance.) pNPIs also have the advantage of being derivable automatically, and might be worth investigating from a linguistic perspective in their own right.

5 Cross-linguistic analysis

Applying our algorithm to English: connections to linguistic typology So far, we have made no assumptions about the language on which our algorithm is applied. A valid question is, does the quality of the results vary with choice of application language? In particular, what happens if we run our algorithm on English?

Note that in some sense, this is a perverse question: the motivation behind our algorithm is the non-existence of a high-quality list of NPIs for the language in question, and English is essentially the only case that does not fit this description. On the other hand, the fact that DLD09 applied their method for extraction of DE operators to English necessitates some form of comparison, for the sake of experimental completeness.

We thus ran our algorithm on the English BLLIP newswire corpus with seed set {'any'} . We observe that, surprisingly, the iterative addition of pNPIs has very little effect: the precisions at k are good at the beginning and stay about the same across iterations (for details see figure 5 in in the externally-available appendices). Thus, on English, co-learning does not hurt performance, which is good news; but unlike in Romanian, it does not lead to improvements.

Why is English 'any' seemingly so "powerful", in contrast to Romanian, where iterating beyond the initial 'any' translations leads to better results? Interestingly, findings from linguistic typology may shed some light on this issue. Haspelmath [9] compares the functions of indefinite pronouns in 40 languages. He shows that English is one of the minority of languages (11 out of 40)[9] in which there exists an indefinite pronoun series that occurs in all (Haspelmath's) classes of DE contexts, and thus can constitute a sufficient seed on its own. In the other languages (including Romanian),[10] no indirect pronoun can serve as a sufficient seed. So, we expect our method to be viable for all languages; while the iterative discovery of pNPIs is not necessary (although neither is it harmful) for the subset of languages for which a sufficient seed exists, such as English, it is essential for the languages for which, like Romanian, 'any'-equivalents do not suffice.

Using translation Another interesting question is whether directly translating DE operators from English is an alternative to our method. First, one should emphasize that there exists no complete list of English DE operators (the largest available collection is the one extracted by DLD09). Second, we do not know whether DE operators in one language translate into DE operators in another language. Even if that were the case, and we somehow had access to ideal translations of DLD09's list, there would still be considerable value in using our method: 14 (39%) of our top 36 highest-ranked Romanian DE operators for iteration 9 do not, according to the Romanian-speaking author, have English equivalents appearing on DLD09's 90-item list. Some examples are: 'abţinut' (abstained), 'criticat' (criticized) and 'reacţionat' (reacted). Therefore, a significant fraction of the DE operators derived by our co-learning algorithm would have been missed by the translation alternative even under ideal conditions.

6 Conclusions

We have introduced the first method for discovering downward-entailing operators that is universally applicable. Previous work on automatically detecting DE operators assumed the existence of a high-quality collection of NPIs, which renders it inapplicable in most languages, where such a resource does not exist. We overcome this limitation by employing a novel *co-learning* approach, and demonstrate its effectiveness on Romanian.

Also, we introduce the concept of *pseudo-NPIs*. Auxiliary experiments described in the externally-available appendices show that pNPIs are actually more effective seeds than a noisy "true" NPI list.

Finally, we noted some cross-linguistic differences in performance, and found an interesting connection between these differences and Haspelmath's [9] characterization of cross-linguistic variation in the occurrence of indefinite pronouns.

[9] English, Ancash Quechua, Basque, Catalan, French, Hindi/Urdu, Irish, Portuguese, Swahili, Swedish, Turkish.

[10] Examples: Chinese, German, Italian, Polish, Serbian.

Acknowledgments We thank Tudor Marian for serving as an annotator, Rada Mihalcea for access to the Romanian newswire corpus, and Claire Cardie, Yejin Choi, Effi Georgala, Mark Liberman, Myle Ott, João Paula Muchado, Stephen Purpura, Mark Yatskar, Ainur Yessenalina, and the anonymous reviewers for their helpful comments. Supported by NSF grant IIS-0910664.

References

[1] Roy Bar-Haim, Jonathan Berant, Ido Dagan, Iddo Greental, Shachar Mirkin, Eyal Shnarch, and Idan Szpektor. Efficient semantic deduction and approximate matching over compact parse forests. In *Proceedings of the Text Analysis Conference (TAC)*, 2008.

[2] Eric Breck. A simple system for detecting non-entailment. In *Proceedings of the Text Analysis Conference (TAC)*, 2009.

[3] Christos Christodoulopoulos. Creating a natural logic inference system with combinatory categorial grammar. Master's thesis, University of Edinburgh, 2008.

[4] Ido Dagan, Oren Glickman, and Bernardo Magnini. The PASCAL Recognising Textual Entailment challenge. In *Machine Learning Challenges, Evaluating Predictive Uncertainty, Visual Object Classification and Recognizing Textual Entailment, First PASCAL Machine Learning Challenges Workshop*, pages 177–190. Springer, 2006.

[5] Cristian Danescu-Niculescu-Mizil, Lillian Lee, and Richard Ducott. Without a 'doubt'? Unsupervised discovery of downward-entailing operators. In *Proceedings of NAACL HLT*, 2009.

[6] David Dowty. The role of negative polarity and concord marking in natural language reasoning. In Mandy Harvey and Lynn Santelmann, editors, *Proceedings of SALT IV*, pages 114–144, 1994.

[7] Gilles Fauconnier. Polarity and the scale principle. In *Proceedings of the Chicago Linguistic Society (CLS)*, pages 188–199, 1975. Reprinted in Javier Gutierrez-Rexach (ed.), *Semantics: Critical Concepts in Linguistics*, 2003.

[8] Bart Geurts and Frans van der Slik. Monotonicity and processing load. *Journal of Semantics*, 22(1):97–117, 2005.

[9] Martin Haspelmath. *Indefinite Pronouns*. Oxford University Press, 2001.

[10] Jack Hoeksema. Corpus study of negative polarity items. *IV-V Jornades de corpus linguistics 1996-1997*, 1997. http://odur.let.rug.nl/~hoeksema/docs/barcelona.html.

[11] Jon Kleinberg. Authoritative sources in a hyperlinked environment. In *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 668–677, 1998. Extended version in *Journal of the ACM*, 46:604–632, 1999.

[12] Wilfried Kürschner. *Studien zur Negation im Deutschen*. Narr, 1983.

[13] William A. Ladusaw. *Polarity Sensitivity as Inherent Scope Relations*. Garland Press, New York, 1980. Ph.D. thesis date 1979.

[14] Timm Lichte and Jan-Philipp Soehn. The retrieval and classification of Negative Polarity Items using statistical profiles. In Sam Featherston and Wolfgang Sternefeld, editors, *Roots: Linguistics in Search of its Evidential Base*, pages 249–266. Mouton de Gruyter, 2007.

[15] Bill MacCartney and Christopher D. Manning. Modeling semantic containment and exclusion in natural language inference. In *Proceedings of COLING*, pages 521–528, 2008.

[16] Rowan Nairn, Cleo Condoravdi, and Lauri Karttunen. Computing relative polarity for textual inference. In *Proceedings of Inference in Computational Semantics (ICoS)*, 2006.

[17] Frank Richter, Janina Radó, and Manfred Sailer. Negative polarity items: Corpus linguistics, semantics, and psycholinguistics: Day 2: Corpus linguistics. Tutorial slides: http://www.sfs.uni-tuebingen.de/~fr/esslli/08/byday/day2/day2-part1.pdf, 2008.

[18] Víctor Sánchez Valencia. *Studies on natural logic and categorial grammar*. PhD thesis, University of Amsterdam, 1991.

[19] Johan van Benthem. *Essays in Logical Semantics*. Reidel, Dordrecht, 1986.

[20] Ton van der Wouden. *Negative contexts: Collocation, polarity and multiple negation*. Routledge, 1997.

7 Appendices

7.1 Labeling protocol

Following DLD09, system outputs (i.e., candidate DE operators) were combined with a set of *distractors* — words related to but not synonymous with the system outputs, according to WordNet. We chose distractors that were similar to real system outputs so that the distractors would not obviously "stand out" as seeming unusual.

The combined collection was presented in randomized order to two Romanian native speakers (one an author, one not), whose task was to label each item as either "DE", "not DE", or "Hard", meaning "hard to decide", and to justify their choice by providing an inference example of the kind we discussed in the Introduction. The judges were informed of the presence of distractors; this served to guard against the judges being biased towards simply defaulting to the "DE" label. We should mention that the annotation process was very time consuming because generating definitive example inferences can be quite difficult,[11] limiting the number of output items that we could get annotated. Also, because creativity can be needed to construct the requisite supporting inferences, the two judges first worked independently, and then conferred to resolve their disagreements. The complete system output and the annotations are publicly available at: www.cs.cornell.edu/~cristian/acl2010/.

7.2 Re-ranking marginal NPIs

In this paper we proposed a DE-operator discovery algorithm that can be applied to languages for which there is no pre-existing collection of NPIs; this is useful because collecting NPIs is, as discussed in the Introduction, quite difficult — determining NPI-hood can be hard even for experts.

However, there are a few languages wherein the effort has been expended to collect a large *noisy* collection of NPIs, and Romanian is one of them. (This motivated our choice of this language.) CoDII-NPI.ro[12] contains 58 Romanian NPIs but in the opinion of the Romanian-speaking author, many are *marginal*, by which we mean that their NPI sense is very infrequent.[13] The existence of CoDII-NPI.ro allows us to ask: which is more effective, a list including both true and marginal NPIs (presuming one is lucky enough to have one), or pseudo-NPIs?

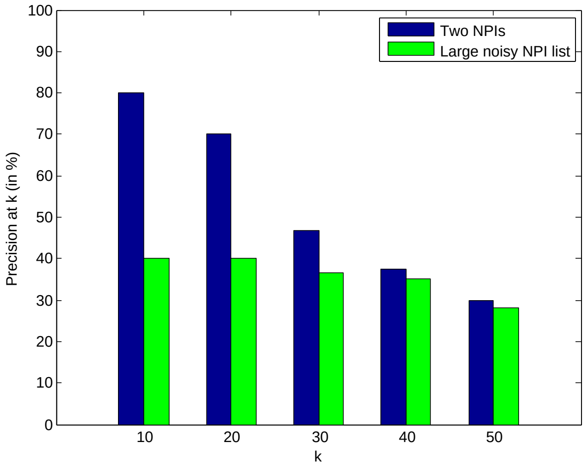

Figure 3: Precision at $k = \{10, 20, ..., 50\}$ using DLD09 on Romanian. NPI lists used: the two translations of 'any' (blue/dark bars); or the relatively large but noisy CoDII-NPI.ro list (green/light bars).

First, it turns out that the CoDII-NPI.ro list containing marginal NPIs is much less effective input to the DLD09 method than are the two 'any'-translation NPIs, as shown in Figure 3. Given that 'vreo' and 'vreun' are contained in CoDII-NPI.ro, we can conclude that the DLD09 algorithm is highly sensitive to the marginal items in that list.

There is another way to employ the CoDII-NPI.ro list: use our co-learning algorithm, but try to have the iterations only add high-quality CoDII-NPI.ro items, rather than arbitrary pseudo-NPIs; essentially, this amounts to *re-ranking* CoDII-NPI.ro. We implement this by altering the pNPI-learning step in our algorithm as follows: extend $\mathcal{N}$ with the top n_r NPIs on the CoDII-NPI.ro list, as opposed to the top n_r pNPIs overall. However, Figure 4 demonstrates that this substantially underperforms the algorithm as originally proposed, i.e., based on pNPIs.

All this suggests that pNPIs might be more effective clues than a mixture of non-marginal and marginal NPIs. In addition, pNPIs are derivable automatically, and might even be worth investigat-

[11] Annotators reported a rate of about 10 examples per hour.

[12] http://www.sfb441.uni-tuebingen.de/a5/codii/, see Trawński and Soehn, "A Multilingual Database of Polarity Items", *LREC* 2008.

[13] For example, CoDII-NPI.ro contains the Romanian verb 'a mişca' ('to move'), which is very frequently used in positive contexts. The inclusion of marginal NPIs is presumably due to a design decision to make the list have high recall.

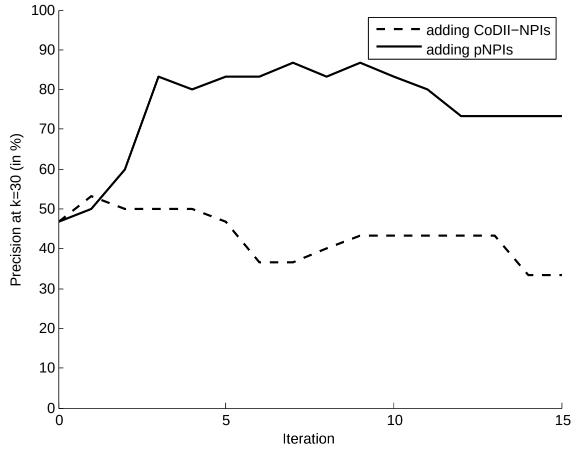

Figure 4: Precision at $k = 30$ obtained by either re-ranking the CoDII-NPI.ro list (dashed line) or by using our co-learning method (solid line). Results for other values of k are similar.

ing from a linguistic perspective in their own right.

7.3 Co-learning results on English

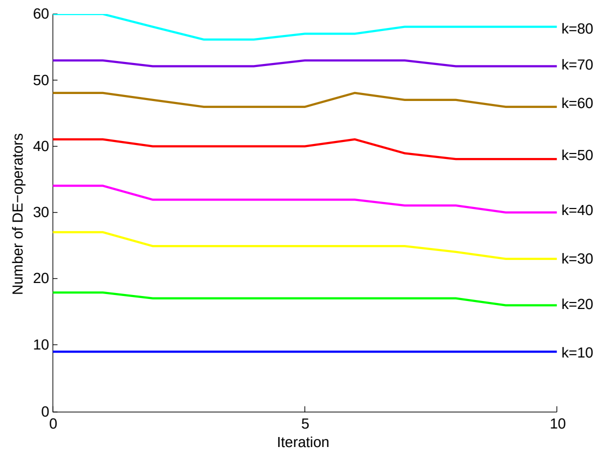

Figure 5: Number of DE operators in the top k results returned by the co-learning method at each iteration for English (seed used: ‘any’).