Volunteers Created the Web

Lucy Vanderwende

Natural Language Processing Group, Microsoft Research
lucyv@microsoft.com

Abstract

This paper shows how the potential for acquiring common sense knowledge from the web, outlined in Singh et al. (2002), can be realized automatically. We describe a system that has been designed to read the web and detect at least some of the kinds of common sense that volunteers have been contributing to Open Mind Common Sense (OMCS). This has at least two significant benefits: by obviating the need to provide certain kinds of knowledge that can be automatically gathered, volunteers can now be directed towards tasks that are less amenable to automation, where their input will be more valuable and effective as a result. Second, it provides a method for broadening the coverage of OMCS knowledge.

Introduction

It is commonly assumed that because every person has common sense, "such knowledge is typically omitted from social communications, such as text" (Liu and Singh 2004, p. 211). This is the intuition that is captured in one of Grice's four maxims of conversation, the maxim of Quantity, "make your contribution as informative as is required for the current purposes of the exchange; do not make your contribution more informative than is required," (as discussed in Levinson 1983, p. 101).

But is common-sense knowledge really absent from everyday communication? Common sense is also "*defeasible*, meaning that it is just a default assumption about the typical case" (Liu and Singh 2004, p. 212, original italics). What of the atypical cases? Shouldn't we expect to find utterances in everyday language, where the default assumption is explicitly overridden? In fact, a study of the literature on presupposition and implicature shows many naturally occurring patterns that explicitly or implicitly indicate that default assumptions are in play. Consider the following example, and its inferences:

1. A bat is the only mammal that can truly fly[1] .
   1a. a bat can fly
   1b. most mammals do not fly

Both (1a) and (1b) are both expressions of common sense, and are very similar to the type of contributions that can be found in Open Mind Common Sense (OMCS).

We propose to mine the text on the Web for expressions of common sense knowledge. We view the authors of the web pages as volunteers, who have contributed common sense information without explicitly intending to do so.

Expressions of common sense

In OMCS, contributors were asked to provide statements of common sense, "all those aspects of the world that we all understand so well we take them for granted" (OMCS home page). The contributors appear to have a fairly good understanding of what constitutes common sense; an initial evaluation of the OMCS knowledge base found that approximately 47% of the statements were judged to be general (e.g., 2 and 3), and 16% judged to be specific facts (e.g., 4) (Singh et al. 2002).

2. Birds often make nests out of grass.
3. Dew is wet.
4. Eritrea is part of Africa.

An open question is whether we can arrive at a definition of the type of knowledge that is being targeted. A second is whether we can arrive at an objective metric for determining whether a statement can be automatically accepted or rejected.

One test could be that common sense is defeasible, and so exceptions must be possible. Indeed, the example in 2 is defeasible. Birds often make nests out of grass, but not always; birds also make nests out of mud, leaves, string, etc. However, the example in 3 is not defeasible; dew is always wet, if it is not wet, it is not dew. Perhaps the example in 3 represents information that is definitional, and perforce true. If we accept that, then we can begin to distinguish between definitional knowledge, common sense knowledge, and general knowledge (for example, that our solar system consists of one moon and 9 planets), as well specific facts for a given domain.

We have not yet attempted such a classification of knowledge types, and so in what follows, we will use a very loose definition of common sense, which appears to include definitional as well as general knowledge.

[1] From: http://members.aol.com/bats4kids2/bird.htm

Common Sense Patterns

There have been many studies on applying patterns to free text in order to extract some type of knowledge (Hearst 1992; Montemagni and Vanderwende 1992; Riloff 1996; Berland and Charniak 1999; Girju, Badulescu and Moldovan 2003; Fleischman, Hovy, and Echihabi 2003; Etzioni et al. 2004). These studies have described using lexical patterns, or a combination of lexical and syntactic patterns; some studies include a machine-learned filtering component to improve the quality of the extracted information.

It is important to note that all of the studies above focus on extracting structured data. For example, Hearst (1992) aimed to extract hyponyms from free text using the lexico-syntactic pattern "such as"; some examples are *hyponym ("bruise", "injury")*, *hyponym ("fruit", "grape")*. Montemagni and Vanderwende (1992) extract several different relation types; some examples are *location-of ("market", "buy")*, *purpose ("cellar", "store")*, and *object-of ("store", "goods")*.

In contrast, the knowledge representation language in OMCS is "plain language"; volunteers have been asked to enter natural language sentences "even a child could understand" (Singh 2002). This is a very exciting and flexible approach to knowledge representation, but it places on the methods described above the additional requirement of generating natural language strings.

Our system uses hand-crafted lexico-syntactic patterns, applied to parse trees, to identify expressions of common sense. It subsequently generates the extracted information as naturally occurring text, as if the text had been hand-written as in the OMCS contributions. We use the generation component for English described in Aikawa et al. (2001), which takes as input a logical form and produces text as output. We can also generate common sense expressions for any of the other languages described in Aikawa et al. (2001), Gamon et al. (2002) or Smets et al. (2003). While we expect that, given enough training data, these hand-crafted patterns could be machine-learned and the relevant stretches of text identified automatically, we believe that the generation component will continue to provide the benefit of producing natural-sounding text that can be used side by side with the text provided by human volunteer contributors.

Targeting those expressions which indicate default assumptions, we have implemented lexico-syntactic heuristics to capture several of the syntactic constructions described in Levinson (1983, p. 181-4) that signal presupposition and implicature, specifically:

- complements of factive verbs
- temporal adverbials
- non-restrictive relative clauses

In English, "know" is a good example of a factive. In the examples below, (a) is the source sentence, and (b) is the common sense sentence(s) generated automatically by our system. We obtained the source sentences in 5 and 6 by issuing the web query "knew that"[2].

5a. The Greeks knew that people located on a high promontory could see considerably farther out to sea than someone low[3].

5b. People located on a high promontory can see considerably farther out to sea than someone low.

6a. The Ancient Egyptians knew that physical trauma could cause injury, and they knew that snake and scorpion bites could cause serious illness[4].

6b1. physical trauma can cause injury

6b2. snake bites can cause serious illness

6b3. scorpion bites can cause serious illness

Temporal adverbials tend to convey reliable background information, whose truth is preserved whether the main clause is true or false. Examples are given in 7 and 8:

7a. These fragments are exploded when gases build up inside a volcano and produce an explosion.[5]

7b1. Gases can build up inside a volcano.

7b2. Gases can produce an explosion inside a volcano.

8a. When the stone was cut, a total of 105 gems were produced, weighing 1063 carats in all.[6]

8b. Stones can be cut.

Non-restrictive relative clauses also can signal background information, as shown in 9:

9a. Dogs, which are usually social and hunt in packs, tend to run down their prey, ….[7]

9b1. Dogs can be usually social.

9b2. Dogs can hunt in packs.

Finally, we also include heuristics that generate common sense expressions from (reduced) restrictive relative clauses. However, instead of extracting general statements of common sense, the restrictive relative clause describes a default assumption for a subset of referents of the concept.

10a. Gases released near the volcano can be measured for changes in quantity and makeup[8].

---

[2] We will provide illustrations from text obtained from the web that are the result of targeted web queries, since we have not yet crawled the web to extract such information. We source each example with the URL in which it occurred.

[3] From: http://www.sfu.ca/philosophy/swartz/flat_earth.htm

[4] http://www.womenintheancientworld.com/medicine%20in%20ancient%20egypt.htm

[5] Article for "volcano" in Encarta, © 1993-2004 Microsoft Corporation. All Rights Reserved.

[6] http://www.a-asystems.com/salvador/htdocs/diamondfacts.html

[7] http://home.globalcrossing.net/~brendel/carniv.html

10b. Some gases are released near the volcano.

11a. Today, industrial-grade synthetic diamonds used as abrasives are produced inexpensively.[9]

11b. Some industrial-grade synthetic diamonds are used as abrasives.

Certainly not all factives, temporal adverbials or relative clauses should be considered as candidates that will contribute statements of common sense, and so there are a number of constraints embodied in the lexico-syntactic rules that identify the sentences in (a) above. For example, the subject of the factive cannot be a personal pronoun; "I know that snake bites cause injury" carries less authority than "the Egyptians knew that snakes bites cause injury". Also, if the statement to be generated contains a demonstrative article (this/that), then it is rejected as being too specific. Furthermore, many generation issues remain, in particular, all involving which fragments of the logical form of the original sentence need to be passed to the generation component in order to reflect accurately what might be inferred from the volunteer's contribution, as well adjusting the tense of the expression to be generated.

In order to mine other forms of common sense expressions from the web, we are experimenting with patterns beyond those suggested by the literature on presupposition. We are currently looking at sentences that express causation directly (shown in sentence 12a). Additionally, we have identified at least one pattern that signals when the default assumption of causation is overridden. We found that, as in sentences like 13a, the phrase "but luckily" conveys a sense of surprise that what had been expected did not in fact occur. We expect that patterns such as "but luckily" can be machine-learned. We continue to work on refining the logical form input to generation; while the sentences in (b) below are understandable, they are not yet fluent English.

12a. However, in recent years there has been an increase in cases of mumps because some parents have chosen not to let their child have the MMR vaccine.[10]

12b. Some parents choosing not to let their child have the MMR vaccine can cause there to be an increase in cases of mumps.

13a. Poor Dad got hurt when a "jumping jack" landed on the back of his neck and he had a nasty burn, which took a while to heal, but luckily didn't leave a scar.[11]

13b. A nasty burn taking a while to heal can cause leaving a scar.

[8] Article for "volcano" in Encarta, © 1993-2004 Microsoft Corporation. All Rights Reserved.

[9] http://www.peel.edu.on.ca/~havenwd/diamond.htm

[10] http://www.unidocs.co.uk/mumps/nhsdirectmumpsleaflet.pdf

[11] Http://www.toowrite.com/toowrite_story.asp?sid=5401

Collectively, the sentences (b) above illustrate the type of common sense items that our system automatically extracts and generates. These items all appear similar in nature to those contributed by volunteers in the OMCS project, encouraging us to view the existing web as a vast, already existing, source of knowledge contributed implicitly by volunteers.

Evaluation of Contributions

Whether the contributions to OMCS are volunteered explicitly, or have been automatically generated based on analysis of free text, as described above, it is important to evaluate their quality.

As Liu and Singh (2004) among others have said, "it is easier to measure coverage and goodness against a system's performance in concrete tasks and applications". In the absence of such an application, and without an operational definition of common sense, it is unlikely that we will achieve strong agreement in the area of evaluation.

Singh et al. (2002) describe functionality in OMCS-2 which would "enable users to judge each other by judging samples of other's knowledge" (Singh et al. 2002, p. 1235). Although we have not yet seen it on-line, exposing this functionality would be of great interest. First, it would allow the study of agreement among contributors in how they judge pieces of knowledge, where perhaps the judges would also be asked to supply the additional information as to whether a common sense item was core or peripheral to the meaning of the concept. Second, we would ideally like to be able to submit common sense expressions automatically extracted by our system to OMCS-2 to be peer reviewed; if the majority of our contributions were judged to be sensible and yielded interesting inferences, the project would be considered a success. Finally, this functionality could be used to make public a measure of importance for each common sense item, and thus, volunteers would have incentive to supply items that receive high marks. Such an incentive mechanism is described in Richardson and Domingos (2003), which incorporates a feedback loop to motivate the users by giving them credit for their input in the context of an application.

In the absence of such functionality, we evaluate the items generated by our system following the methodology of Singh et al. (2002). In that paper, the authors describe an evaluation of the quality of the volunteer contributions in OMCS. Human judges were asked to rate the common sense statements on their generality, truth, neutrality, and sense.

We conducted our evaluation, using the same methodology, but with only one judge (not the author). We consider this a preliminary evaluation only, since we cannot measure inter-annotator agreement. As we have not yet crawled any portion of the web, the corpus that we used is generated automatically by our system from the dictionary definitions from the *Longman Dictionary of Contemporary English* and the *American Heritage*

Dictionary, 3rd Edition; our patterns do not extract definitional information, but rather extract those segments of text that convey common sense information. 300 items were chosen at random.

In the evaluation, 19.6% of the OMCS test data and 26% of the automatically generated test data was considered nonstandard or garbage, and discarded.

	Generality	Truth	Neutrality	Makes Sense
Human	3.26	4.28	4.42	4.55
Automatic	2.4	4.83	4.86	4.66

Figure 1. Average rating for each category. Judgements are on a scale from 1 to 5, where 1 is low and 5 high.

The average scores for truth and neutrality are high, which is not surprising given that the data is generated from definitions, which have been written to be truthful and not opinionated. We are encouraged that there is a high average score for the category "makes sense" (92% of the items were rated 4 and higher); this indicates that the patterns target interesting and reasonable information, although primarily the pattern that identifies restrictive relative clauses is used, as can be seen in the examples given below. The average score for generality is lower than hoped for; we expect that data generated from less specialized text will show better results.

Examples of statements scoring high on "sense" are:

- Some birds walk in the water.[12]
- Some people are ready to fight or struggle.[13]

And some statements scoring low on "sense" are:

- Some people stay away[14]
- Some expressions ask a question[15]

Examples of statements scoring low on generality are:

- Some brown seeds are rich in veratrine.[16]
- Some large, inedible roots yield an extract used as a raw material for synthetic steroid hormones.[17]

Examples of statements scoring high on generality are:

- Some obstacles must be overcome.[18]
- Some people take great interest in the pleasures of food and drink.[19]

More examples for each of the categories are included in the appendix.

Despite these encouraging first results, there are two areas that are of concern in this evaluation methodology. First, the question raised earlier in this paper remains; in the absence of an operational definition of common sense, the results of any evaluation will always be less than satisfactory. Casual inspection of sample contents from OMCS, for example, always reveals several items which convey definitional information, such as "cobras are venomous snakes". And casual inspection of sample content extracted from our dictionary corpus includes several items which would be considered general information, and not common sense, such as "some bicycles can be pedaled in ten different gears". Until it is clear what it is we are measuring, or what type of information is required for a particular application, it is difficult to present meaningful evaluation results for common sense information – regardless of how it is gathered.

Measuring Salience of Contributions

In the evaluation of ConceptNet (Liu and Singh 2004), humans are asked to provide a judgment of exhaustiveness. The authors are encouraged by the results that indicate that "with regard to comprehensiveness, ConceptNet's concepts were judged as containing, on average several relevant concepts, but varied significantly from *a few concepts* to *almost all of the concepts*" (Liu and Singh 2004, p. 223, original italics).

Our concern with evaluating exhaustiveness in this way is that it conflates recall and importance. In fact, not all relevant concepts are equally important. Suppose there are three relevant concepts associated with the concept "bird": "birds fly", "birds swim" and "birds molt". If all relevant concepts were equally important, then the judge would give a high rating only if all three concepts were present. Consider, however, the following two scenarios: first, the judge is presented with "birds fly". Presumably, the judge will rate this high, since the most salient relevant concept

[12] From the definition of "*water bird*", "any bird that swims or walks in the water"
[13] From the definition of *warpath*, "the course of action of a person who is ready to fight or struggle"
[14] From the definition of *absentee*, "a person who stays away"
[15] From the definition of *interrogative*, "a sentence or an expression that asks a question"
[16] From the definition of *sabadilla*, "a Mexican and Central American plant "(Schoenocaulon officinale)" of the lily family, having ..., and brown seeds that are rich in veratrine"
[17] From the definition of *barbasco*, "any of several Mexican plants of the genus "Dioscorea" having a large, inedible root that yields an extract used as a raw material for synthetic steroid hormones."
[18] From the definition of "*obstacle course*", "a situation full of obstacles that must be overcome"
[19] From the definition of *epicure*, "a person who takes great interest in the pleasures of food and drink, and regards cooking as an art; gourmet"

for “bird” is presented, namely “fly”. In the second case, the judge is presented with “birds molt” and “birds swim”. Undoubtedly, the judge will rate this lower, since the most frequent related concept is missing, even though in this case more relevant concepts are present.

It would be interesting to redesign this evaluation to incorporate a measure of salience of common sense items; this would also provide a sense of utility, which participants wished to have (Singh et al. 2002, p. 1229). As suggested in Etzioni et al. (2004) and Soderland et al. (2004), we propose to use web statistics to validate the salience of the extracted information, specifically, we propose computing pointwise mutual information (PMI) from hit counts (see Turney 2001). Intuitively, the formula below computes the probability that C1 and C2 will be found together on a web page that contains C1.

PMI(C1, C2) = Hits(C1 + C2) / Hits(C1)

Using the search engine Google on 1/25/05[20], the following PMI scores were obtained for the expressions relating to “bird” (C1):

C1 + C2	Hits	PMI
bird + fly	1,740,000	0.139
bird + swim	365,000	0.029
bird + molt	26,600	0.002
bird	12,500,000	

In this case, PMI demonstrates that the importance of capturing information about “birds fly” is an order of magnitude greater than the importance of “birds swim”, which in turn is an order of magnitude greater than the importance of “birds molt”. Using web counts to compute the measure of salience is another way to leverage the information that web page authors have collectively contributed.

Just as such a measure would allow one to begin to answer whether volunteers contribute salient and important information, this measure can be used to evaluate items that have been automatically generated by our system. For example, having processed the information from several web sites on volcanoes, our system generated the following expressions, provided in descending order of PMI score:

- Volcanoes are hot. (0.16)
- Volcanoes can erupt. (0.02)
- Gases can build up inside a volcano. (0.01)
- Some gases are released near the volcano. (0.006)

It has also been suggested that web statistics can be used to make a distinction between what is common sense information and what is general information. One might assume that if a particular expression is rarely expressed overtly, and instead, is predominantly found expressed as a default being overridden, that it is likely to be common sense. Indeed, in our project, we have seen very few instances of common sense items that were generated by our system, and found verbatim on the web. To explore this further, we need to devise tests which would categorize statements as various types of information.

[20] We used the citation forms of the words.

Expanding a Knowledge Base

An inherent problem for knowledge bases constructed by handcrafting data (including Cyc (Lenat 1995), WordNet (Fellbaum 1998), and ConceptNet (Liu and Singh 2004)), is that all will suffer from gaps in their data. It is possible that nothing about a requested concept will be available; this was the case in the hand evaluation of ConceptNet, where approximately one out of ten concepts that the human evaluators searched for was missing (Liu and Singh 2004). Or, some of the concept information was present, but perhaps not all of the salient information needed; this too was shown to be the case in the ConceptNet evaluation, where comprehensiveness for the concepts requested averaged 3.4 on a scale from 1 to 5 (where 5 is the highest score given). This problem is not specific to ConceptNet – systems that make use of WordNet data often comment that their coverage is less than expected because of gaps in WordNet data. Applications that use common sense knowledge, then, need to have a mechanism for acquiring the necessary information on demand.

Because our system acquires and generates common sense items automatically, it can be used to fill in the gaps in any existing knowledge base. The salience measure suggested above, PMI, can be used to discover areas where salient information is currently missing. If an external resource provides the set of words most likely to co-occur with a specific concept, and no relation is found between those words in the current knowledge base, then a search engine can be directed to retrieve passages containing both words in the word pair. These passages can then be processed using our system to discover any common sense relations holding between the concepts.

For example, one of the volunteer contributions to OMCS is that the purpose of “rope” is “to hang people”; we calculate a PMI of 0.08 for “rope hang”. If a system needs information about “rope”, are the applications of OMCS compromised if the more salient purposes of “rope” are missing? We calculate that the PMI of “rope hold”, for example, is 0.15. If so, word co-occurence, directed search queries, and automatic identification and extraction of items are a means to supplement the missing information automatically.

Conclusion

This paper describes an automated method for identifying and generating common sense expressions. These items were contributed implicitly, by volunteers constructing their web pages on a multitude of topics. By analyzing the already existing text, many interesting common sense items can be identified and generated, which can supplement the common sense knowledge base, OMCS.

Since at least some types of common sense can be generated automatically, and since definitional information can be also be reliably generated, we suggest that the efforts of volunteer contributors can be directed to contributing knowledge not captured explicitly in the surface constructions of natural language. From the point of improving natural language understanding, some of the most exciting activities volunteers can contribute to are "cause and effect", "what changed", and, especially, "what else should I know".

Finally, it is essential for any knowledge base to be able to grow dynamically; the knowledge bases that have relied on handcrafting data, Cyc, WordNet, and ConceptNet included, all will suffer from gaps in their data that need to be filled on demand.

Acknowledgements

I would like to thank my colleagues, Arul Menezes, Bill Dolan and Simon Corston-Oliver, and other members of the Natural Language Processing group at Microsoft Research. I especially would like to thank Deborah Coughlin for help with evaluation. I would like to thank the reviewers for their interesting comments, many of which remain to be addressed, and finally to Rada Mihalcea for encouraging me to write this paper.

Appendix

All of the examples below are generated from definitions in either the *Longman Dictionary of Contemporary English* or the *American Heritage Dictionary, 3rd Edition*.

Examples judged to be high in Generality:

- Some rivers are not wide.
- Some people live on the beach or near the beach.
- Some blades are sharp on the outer edge.
- Some sticky substances are used for joining things together.
- Some people try to make others laugh by doing foolish things.

Examples judged to be low in Generality:

- Some thermosetting resins are used for molded products, adhesives, and surface coatings.
- Some retail stores stock merchandise on consignment.
- Some orange flowers produce seeds containing oil used in cooking, cosmetics, paint, and medicine.

Examples judged to be high in Truth:

- Some very wrong or cruel acts cause great anger.
- Some submarines attack a single vessel or convoy.
- Some practice attempts are made before the real thing.

Examples judged to be low in Truth:

- Pregnancy poses a danger to the woman's health
- Some airplane wings reduce lift

Examples judged to be high in Neutrality:

- Some retail stores stock merchandise on consignment.
- Some fragrant roots are used as a flavoring.
- Some balls curve away from the batter.

Examples judged to be low in Neutrality:

- Some small guitars are used in playing non-serious music.
- Some people must be paid back at an unfairly high rate of interest.
- Some homeless people live as a derelict.

Examples judged to be high in "Makes Sense":

- Some substances induce sleep.
- Some cosmetics are used as a base for facial makeup.
- Some blades are sharp on the outer edge.

Examples judged to be low in "Makes Sense":

- Some assignments are undertaken.
- Some long roots are eaten raw.
- Some cases are put on the floor for sleeping on.

References

Aikawa, T.; Melero, M.; Schwartz, L.; and Wu, A. 2001. Multilingual Sentence Generation. In Proceedings of 8th European Workshop on Natural Language Generation, Toulouse, France.

Berland, M. and Charniak, E. 1999. Finding parts in very large corpora. In Proceedings of the North American ACL, 57—64. College Park, MD.

Etzioni, O.; Cafarella, M.; Downey, D.; Popescu, A-M.; Shaked, T.; Soderland, S.; Weld, D.; and Yates, A.. 2004. Methods for Domain-Independent Information Extraction from the Web: An Experimental Comparison. In AAAI 2004, San Jose, CA.

Fellbaum, C. (ed). 1998. WordNet: An Electronic Lexical Database. MIT Press.

Fleischman, M.; Hovy, E; and Echihabi, A. 2003. Offline Strategies for Online Question Answering: Answering Questions Before They Are Asked. In Proceedings of the seventh Annual Meeting of the Association for Computational Linguistics, 1-7. Sapporo, Japan.

Gamon, M., E. Ringger, and S. Corston-Oliver. 2002. Amalgam: A machine-learned generation module. Microsoft Research Technical Report: MSR-TR-2002-57.

Girju, R.; Badulescu, A.; and Moldovan, D. 2003. Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations. In Proceedings of the Human Language Technology Conference (HLT-NAACL), 1-8. Edmonton, Canada.

Hearst, M. 1992. Automatic acquisition of hyponyms from large text corpora. In Proceedings of the Fifteenth International Conference on Computational Linguistics, 539-545. Nantes, France.

Lenat, D. 1995. CYC: A large-scale investment in knowledge infrastructure. In Communications of the ACM, 38, No 11. 33-38.

Levinson, S. 1983. Pragmatics. Cambridge University Press.

Liu, H., and Singh, P. 2004. ConceptNet: a practical commonsense reasoning toolkit. BT Technology Journal, 22(4):211-226.

Montemagni, S., and Vanderwende, L. 1992. Structural patterns vs. string patterns for extracting semantic information from dictionaries. In Proceedings of the Fifteenth International Conference on Computational Linguistics, 546-552. Nantes, France.

Richardson, M, and Domingos, P. 2003. Building Large Knowledge Bases by Mass Collaboration. In Proceedings of the Second International Conference on Knowledge Capture, 129-137. Sanibel Island, FL.

Riloff, E. 1996. Automatically Generating Extracting Patterns from Untagged Text. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, 1044-1049. The AAAI Press/MIT Press.

Singh, P. 2002. The Public Acquisition of Commonsense Knowledge. In Proceedings of AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access. Palo Alto, CA.

Singh, P.; Lin, T.; Mueller, E.; Lim, G.; Perkins, T.; and Zhu, W.L. 2002. Open Mind Common Sense: Knowledge acquisition from the general public. In Robert Meersman & Zahir Tari (Eds.), Lecture Notes in Computer Science: Vol. 2519. On the Move to Meaningful Internet Systems 2002: DOA/CoopIS/ODBASE 2002. 1223-1237. Heidelberg: Springer-Verlag.

Smets, M., M. Gamon, S. Corston-Oliver and E. Ringger. 2003. The adaptation of a machine-learned sentence realization to French. In Proceedings of the 10th conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary, pp. 323-330.

Soderland, S.; Etzioni, O.; Shaked, T.; and Weld, D. 2004. The Use of Web-based Statistics to Validate Information Extraction. In AAAI 2004 Workshop on Adaptive Text Extraction and Mining, San Jose, CA.

Turney, P. 2001. Mining the Web for synonyms: PIM-IR versus LSA on TOEFL. In Proceedings of the Twelfth European Conference on Machine Learning.