

WordNet Hierarchy Axiomatization and the Mass–Count Distinction

Jonathan Gordon
Department of Computer Science
University of Rochester
Rochester, New York, USA
Email: jgordon@cs.rochester.edu

Lenhart K. Schubert
Department of Computer Science
University of Rochester
Rochester, New York, USA
Email: schubert@cs.rochester.edu

Abstract—WordNet provides a semantic hierarchy with broad lexical coverage, which has proved sufficiently precise to boost performance at many tasks involving natural language. However, it has not yet been formalized for use in a general reasoning system. In this paper, we present such a formalization. We use a semi-automatic annotation of WordNet with lexical features – most notably the mass–count distinction – to recognize inferentially different relations between concepts. The result is a collection of 77,263 lexical-semantic axioms, which are being released for general use. We evaluate a sample of the axioms for core concepts, showing their quality to be significantly better than a baseline interpretation of WordNet.

I. INTRODUCTION

We are interested in the creation of large knowledge bases to support language understanding and commonsense problem-solving. An important component of such a knowledge base is a large collection of lexical-semantic axioms relating more specific nominal concepts to more general ones. For example, axioms might assert that every rifle is a firearm, or that all malpractice is wrongful conduct. The most comprehensive, machine-readable source of this type of knowledge is WordNet [1], which attempts to exhaustively enumerate and define the senses of each word. WordNet groups word senses considered synonymous into *synsets*, such as {firearm#1, piece#7, small-arm#1} or {wrongdoing#2, wrongful_conduct#1, misconduct#2, actus_reus#1}. These synsets are linked by hierarchy relations: upward to a more general *hypernym* and downward to a more specific *hyponym*.¹

These relations have been used to improve performance in natural language processing tasks such as information retrieval, document clustering, rule discovery, and text-based question-answering. These incremental improvements have not required that the knowledge used be sufficiently precise to support genuine language understanding, let alone commonsense reasoning. For these “deep” problems, the knowledge needs to be regimented into some more precise, more reliable form. At the same time, it is desirable to keep a close connection between concepts referred to in ordinary language and their formalized versions. This way, the mapping from natural

language to formalized representations (and vice versa) will be as direct and straightforward as possible in such applications as text-understanding and human–computer dialogue.

As such, it is natural to ask whether WordNet senses of nouns can be used directly as predicates in a formalized knowledge base, with hierarchy relations (in the upward direction) corresponding to universally quantified conditional (i.e., if–then) formulas. For example, can we make the formal claim that, for appropriate senses, every rifle is a firearm, or that all malpractice is wrongful conduct, as in the following?

$$\begin{aligned}\forall x. \text{rifle}_{1.n}(x) &\Rightarrow \text{firearm}_{1.n}(x), \\ \forall x. \text{malpractice}_{1.n}(x) &\Rightarrow \text{wrongful_conduct}_{1.n}(x)\end{aligned}$$

For people to judge these claims, we want to verbalize them as corresponding English statements:

Every rifle is a firearm.
Every amount of malpractice is an amount of wrongful conduct.

The difference in these phrasings reflects a difference in the meaning of the terms involved: Rifles are distinct, individual entities; malpractice is less discrete. More generally, while mass terms are cumulative, count terms are not [2]. A test phrasing highlights the difference:

Some malpractice and some more malpractice constitute an amount of malpractice.
*Some rifle and some more rifle constitute an amount of rifle.

Consistent with the close semantic connection between mass terms and plurals [3], if we change ‘rifle’ to ‘rifles’, the claim becomes true, even if speakers disprefer the phrasing ‘amount of’ applied to a plural.

For other hyponym–hypernym pairs, different relations are appropriate. For instance, the count noun ‘plank’ and the mass noun ‘lumber’ are related as *Every plank is an amount of lumber*. When a hyponym denotes an individual or a generic kind, the appropriate relation is not subsumption but instantiation, e.g., *Gold is a noble metal*. We sort out this ambiguity by looking at logical test phrasings and considering the criteria for detecting these meanings, most notably the mass–count distinction. In Section IV, we enumerate the relations we find between synset members in WordNet’s hypernym hierarchy

¹WordNet also includes antonymy, part–whole, and membership relations among others. In this paper we focus on nominal hypernyms, WordNet’s most extensive and most used component.

and the conditions under which they hold, but we first consider whether some alternative lexical and ontological resources would be less problematic and whether previous work in refining WordNet could help.

II. PREVIOUS WORK

One of the most noteworthy efforts in knowledge engineering for artificial intelligence is the Cyc project [4], which has created a collection of world knowledge in the CycL logical form, including a manually constructed core ontology of over 200,000 terms. However, Cyc lacks the systematic link to language we find in WordNet, as seen in constructed predicates like *CoexistingWithSomethingElse* or *Organism-ByTaxonomicKingdom-Biology-Topic*.

Another significant resource is SUMO [5], a formal ontology consisting of around 80,000 axioms, including information about classes and their meanings. Álvarez et al. [6] translated most of its 1,000-term upper ontology into first-order logic (FOL) and demonstrated its use for commonsense inference. However, while the logical formulation of axioms in SUMO is appealing for reasoning, it does not meet our goals of lexical and conceptual coverage. SUMO maps WordNet synsets to its own formal terms for broader coverage, but these mappings are coarse and lose the specific meaning of the synset. For instance, in WordNet, *stage dancing* (e.g., ballet) has the hypernyms *dancing* (the act) and *performing arts* (the discipline). In SUMO, ‘stage dancing’ is mapped to the class *Dancing*, but there is no sense for this as a discipline or an art, only as *BodyMotion*.

Pustejovsky [7] presents the alternative approach of a “generative lexicon”, recognizing that in different contexts a word will express different meanings, making it infeasible to try (as WordNet does) to enumerate them independently of context. Rather, he argues, a lexical entry should provide the information necessary to derive the sense the word will take on in a given context. Resulting work on the Brandeis Semantic Ontology [8] may eventually provide a more consistent basis for lexical axioms, but no resource has yet been released.

For WordNet, Kaplan & Schubert [9] previously looked at the accuracy of taking the noun hierarchy as a simple subsumption taxonomy. They identified a number of the problems we address in this paper, including the conflation of individuals, predicates, and kinds and the mixing of mass and count uses of terms. More recently, Verdezoto & Vieu [10] presented promising work to automatically identify problematic relations in WordNet based on conflicts between meronyms and hyponyms. However, neither effort attempted to produce a corrected formal resource as we do here.

Several lines of previous work have sought to address “is-a” ambiguity by making WordNet a formal ontology, manually separating or removing non-subsumptive hypernym relations and restructuring the upper, most abstract levels of the hierarchy to fit different ontological principles. Most notable is the work of Nicola Guarino and his collaborators (e.g., [11])

on distinctions and design principles for producing cleaner ontologies. This resulted in the construction of the DOLCE upper-level ontology and the alignment of WordNet (1.6) subtrees to it, to make OntoWordNet [12].

In contrast with these lines of work, we are less concerned with ontological hygiene than with inferential efficacy for intuitively plausible reasoning and understanding. Rather than restrict the content of WordNet to its subsumptive relations, we automatically produce lexical axioms that formalize a variety of logical relations between synset members. To the best of our knowledge, this is the first such attempt.

III. KNOWLEDGE REPRESENTATION

The axioms we produce are in Episodic Logic (EL), a highly expressive formal representation, well suited to reasoning about language. For readability, EL uses infix form for predication, e.g., [Joe.name happy.a], [Joe.name love.v Sue.name] rather than happy.a(Joe.name), love.v(Joe.name, Sue.name). This also distinguishes predication from the prefix form used for function application, e.g., (father-of John.name). EL is language-like, and this is reflected in mnemonic extensions such as ‘name’, ‘n’, and ‘v’ for individual constants, noun-like predicates, and verb-like predicates, respectively. More importantly, EL allows for the expressive devices that are found in all natural languages, including generalized quantifiers (e.g., *all or most*), predicate and sentence modifiers (e.g., *very, perhaps*), and predicate and sentence reification operators (e.g., *that*).

EL and its inference engine, EPILOG, have proved their versatility and effectiveness in experimental applications ranging from processing aircraft maintenance reports to reasoning about fairy-tales. Recent work has even demonstrated self-awareness and autocognitive reasoning – i.e., reasoning dependent on an agent’s knowledge about its own knowledge and its perceptual and cognitive characteristics. See, for example, [13]–[15]. While these experimental applications have been on a small scale, EPILOG has also been shown to hold its own against state-of-the-art systems in the area of shallow theorem-proving over significantly large first-order knowledge bases [16], despite the fact that it handles a much richer representation than FOL. There is also a close connection between the EL/EPILOG approach to commonsense reasoning and “natural logic” [17], which in recent years has proven effective for simple entailment inferences and shows the advantages of staying close to language in building commonsense inference machinery.

IV. ACQUIRING AXIOMS

No work has yet provided a large-scale set of reasonably reliable lexical axioms that are closely integrated with language. To address this problem, we studied random samples of WordNet hypernym relations, formulated appropriate logical axioms, and then formed hypotheses about what features indicate that a relationship holds. In this section, we present

the types of relations we find in WordNet, the method of generating axioms, and the criteria we use.

A. Distinguishing Relations

In Table I, we give the frequency of axioms resulting from the following schemata, both in WordNet as a whole and in the “core” axioms used for evaluation (see Section V). The axiom schemata are presented in conceptual groups, which are used to balance the evaluation sample.

Group 1: Count Nominals

1 *Every φ is a ψ .*

($\forall x: [x \varphi.n] [x \psi.n]$)

φ and ψ are singular count nominals. The hypernym relation holds for each individual, e.g.,

Every cat is a feline.

($\forall x: [x \text{cat1.n}] [x \text{feline1.n}]$)

Group 2: Mass Nominals and Plurals

2 *Every amount of φ is an amount of ψ .*

($\forall x: [x \varphi.n] [x \psi.n]$)

φ and ψ are either mass terms (‘water’) or lexical plurals (‘cattle’). E.g.,

Every amount of red wine is an amount of wine.

($\forall x: [x \text{red_wine1.n}] [x \text{wine1.n}]$)

This shares the logical form of Schema 1 but differs in its verbalization. As discussed in the introduction, this is an indication that singular count nouns are inferentially distinct from plurals and mass terms with respect to cumulativity. See the inference example in Section VI.

Group 3: Kinds and Individuals

3 *(The) φ is/are a ψ .*

[$\varphi.\text{name } \psi.n$]

φ is an individual – either an individual name (‘Belgium’) or a name-like designation (‘Homo sapiens’) for a generic kind [18]. ψ is a singular count nominal. E.g.,

AIDS is an immunodeficiency.

[AIDS.name immunodeficiency1.n]

4 *(The) φ is/are a ψ .*

[(k $\varphi.n$) $\psi.n$]

φ is a mass nominal (‘oil’) or a count nominal known to function as a natural kind (‘tiger’). ψ is a singular count nominal that is a kind-level predicate (‘species’). This schema gives claims about generic kinds, formed with EL’s kind reification operator ‘k’. E.g.,

Gold is a noble metal.

[(k gold3.n) noble_metal1.n]

5 *Every φ is an item of ψ .*

($\forall x: [x \varphi.n] [x \text{item-of.n (k } \psi.n)]$)

φ is a singular count nominal. ψ is an *atomic ensemble*, a mass term that cannot be arbitrarily subdivided (‘furniture’, not ‘water’); see Section IV-B. Equivalently, ψ

can be a plural; the atomic ensembles denoted by plurals are in no way logically distinguishable from atomic ensembles denoted by mass terms. E.g.,

Every bomb is an item of weaponry.

($\forall x: [x \text{bomb1.n}] [x \text{item-of.n (k weaponry1.n)]$)

6 *(The) φ is/are a branch of (the) ψ .*

[(k $\varphi.n$) branch-of.n (k $\psi.n$)]

φ and ψ are fields of study. These are hyponym descendants of discipline#1, excluding kind-level predicates like *humanistic_discipline#1*. E.g.,

Astronomy is a branch of physics.

[(k astronomy1.n) branch-of.n (k physics1.n)]

Group 4: Transitional

7 *Every φ is an amount of ψ .*

($\forall x: [x \varphi.n] [x \psi.n]$)

φ is a singular count nominal and ψ is a plural or a mass noun. E.g.,

Every document is an amount of written material.

($\forall x: [x \text{document1.n}] [x \text{written_material1.n}]$)

8 *Every amount of φ is a ψ .*

($\forall x: [x \varphi.n] [x \psi.n]$)

φ is a mass term or a plural that is a hyponym of group#1 or measure#2. ψ is a singular count nominal. E.g.,

Every amount of people is a group.

($\forall x: [x \text{people1.n}] [x \text{group1.n}]$)

9 *Every amount of φ is an amount of ψ s.*

($\forall x: [x \varphi.n] [x (\text{plur } \psi.n)]$)

φ is a mass term or a plural count nominal, and ψ is a singular, object-level predicate. E.g.,

Every amount of baggage is an amount of cases.

($\forall x: [x \text{baggage1.n}] [x (\text{plur case5.n})]$)

Group 5: Events

10 *Every φ is a ψ .*

($\forall x: [x (\text{plur } \varphi\text{-c.n})] [x (\text{plur } \psi\text{-c.n})]$)

φ and ψ are events that have both a mass and a count sense. Here the ‘plur’ operator is “massifying” the count (discrete-only) sense of an event predicate to match possible iteration. E.g.,

Every restoration is a repair.

($\forall x: [x (\text{plur restoration2-c.n})] [x (\text{plur repair1-c.n})]$)

11 *Every φ is a ψ .*

($\forall x: [x (\text{plur } \varphi\text{-c.n})] [x \psi.n]$)

φ and ψ are events. φ is ambiguous between mass and count senses, while ψ is count. E.g.,

Every sinning is a transgression.

($\forall x: [x (\text{plur sinning1-c.n})] [x \text{transgression1.n}]$)

12 Every φ is a ψ .

($\forall x: [x \varphi.n] [x (\text{plur } \psi\text{-c.n})]$)

φ is count and ψ is ambiguous between mass and count senses. E.g.,

Every dance step is a locomotion.

($\forall x: [x \text{dance_step1.n}] [x \text{locomotion2-c.n}]$)

B. Method

While the members of each synset are closely related, we found they are not always interchangeable as predicates. Many synsets contain a mix of mass and count, singular and plural, e.g., {cutlery#2, eating_utensil#1}. An eating utensil is an *item of* cutlery. Thus, in relating the synset to its hypernym, tableware#1, we form separate axioms for both of these predicates. However, making an axiom for every combination of word senses in a pair of synsets would lead to an unnecessary explosion in the number of axioms. Instead, those members that share logically equivalent properties (mass terms, atomic ensembles & lexical plurals, singular count nouns, individual names) can be stated to be synonymous and an axiom can use a single representative predicate.

Thus, our method is: For each hyponym–hypernym pair, select the synset members with distinct properties for which we will form axioms. Then for the Cartesian product of the selection sets, check each pair of word senses against the restrictions for each schema and output an axiom when they match:

S , the set of nominal synsets in WordNet
 \mathcal{H} , the set of all hyponym–hypernym synset pairs
 \mathcal{C} , the set of all annotation categories.

GENERATE-AXIOMS(\mathcal{H}):

```

for synset pair  $\langle P, Q \rangle \in \mathcal{H}$ :
   $L \leftarrow \text{SELECT-MEMBERS}(P)$ 
   $L' \leftarrow \text{SELECT-MEMBERS}(Q)$ 
  for lemma pair  $\langle \varphi, \psi \rangle \in L \times L'$ :
    for each schema  $s$ :
      if  $\langle \varphi, \psi \rangle$  matches the argument restrictions of  $s$ :
        Instantiate  $s$  with  $\langle \varphi, \psi \rangle$ 

```

SELECT-MEMBERS($s \in S$):

```

 $R \leftarrow \{\}$ 
for lemma  $l \in s$ :
  for category  $c \in \mathcal{C}$ :
    if  $l$  is annotated as  $c$ :
       $f \leftarrow \text{False}$ 
      for  $r \in R$ :
        if  $r$  is annotated as  $c$ :
          if  $l$  has a lower sense number than  $r$ :
            Replace  $r$  with  $l$  in  $R$ 
       $f \leftarrow \text{True}$ 
      break
    if not  $f$ :
       $R \leftarrow R \cup \{l\}$ 
return  $R$ 

```

C. Determining the Mass–Count Distinction and Other Significant Features

Annotating WordNet with mass and count information requires not just determining if a word is usually mass or

Table I
 AXIOM COUNTS FOR WORDNET AS A WHOLE AND ‘CORE’ SYNSETS

Schema	All Axioms	Core Axioms
1	47,450	4,397
2	7,114	979
3	2,838	46
4	1,998	497
5	887	119
6	637	23
7	5,233	578
8	187	23
9	4,757	430
10	1,892	649
11	1,606	452
12	2,664	553
All	77,263	8,746

count but, for many ambiguous words, whether a particular word sense is. While sometimes WordNet splits a word into different synsets for its mass and count senses, in other cases it coerces a single synset’s meaning through multiple hypernym links. For instance, coffee#1 is both a liquid#1 (mass) and a beverage#1 (count), reflecting different uses:

‘How much coffee did you drink?’

‘I’ll have a coffee.’

Often we consider one form to be basic and the other derived. While coffee is primarily mass, we understand a count use to mean a standard portion of it, i.e., a cup. (See comments in Section VIII.)

Various past studies have been aimed at classifying lexemes as mass, count, or both, e.g., [19]–[23]. Typically these have used multiple sources of information, such as morphology, corpus occurrence environments, the Cyc knowledge base, and seemingly similar lexemes in WordNet. While Álvarez et al. [24] semi-automatically annotated WordNet 1.6 with EuroWordNet’s Top Concept ontology semantic features, including *Substance* and *Object* – rough analogues of mass and count – we found these annotations too noisy, e.g., labeling *cystostome* (a cell mouth) a substance.

We annotate each general noun sense in WordNet 3.1 as a plural or singular count term, an atomic ensemble, or a non-atomic mass term. To do so, we matched patterns for standard mass and count syntactic environments (based on those discussed by Bunt [25]) against the Google n-grams data set [26]. While this gives us broad lexical coverage, it only yields moderate accuracy. E.g., frequent occurrence of ‘a x ’ is a good indication that x has a count sense. However, it is easy to erroneously match references such as ‘Grade A milk’. We supplement this classification by looking up each lemma online in the Oxford Dictionary (<http://oxforddictionaries.com>) and Wiktionary (<http://wiktionary.org>), both of which identify some of their noun senses as mass or count.

When, based on this information, a lemma is ambiguous between mass and count readings, we consider features of the synset it occurs in: the definition and the examples, its

hyponyms and hypernyms, and the other lemmas in the synset. For instance, if the synset definition begins with ‘material...’, it is likely a mass sense. If a synset x has a hyponym whose gloss begins ‘an x that...’, then x is probably a count sense. If the word is ambiguous but the other members of the synset are all known to be mass, then it is probably the mass sense.

In axiom schema 5, we also need to recognize those mass predicates that apply to *atomic ensembles* (also called *aggregate terms*). Unless we’re thinking scientifically, some mass nouns can be divided arbitrarily into more of the same: All *air* has proper parts that are also air, all *meat* has proper parts that are also meat, etc. On the other hand, it is not the case that all *poultry*, *furniture*, *foliage*, *cutlery*, or *dinnerware* has proper parts that are also poultry, furniture, etc. respectively. Rather, these mass terms denote entities that have atomic parts. E.g., a chair is furniture, but it has no proper parts that are also furniture. We are not aware of any attempt to enumerate atomic ensembles in English or to automatically find them in text. For this work, we annotated 178 WordNet lemmas as atomic ensembles with reference to the linguistic literature. An additional 915 lemmas are identified (based on dictionary entries) as lexical plurals. As we noted for axiom schema 5, plurals also denote atomic ensembles, which are logically indistinguishable from mass ones.

It is also important for us to distinguish individuals (e.g., ‘Nikola Tesla’) from common nouns. Following the criticism of ontologists like Aldo Gangemi, WordNet 2.1 began to move instances from hyponym relations to instantiation relations. In WordNet 3.1, there are 15,675 such word senses. These give us a source for formulas about many important events, people, states, and other kinds of individuals. Identifying individuals also lets us avoid nonsensical quantification over ‘Every Nikola Tesla’. However, we find that many individuals are still mingled with classes as hyponyms, e.g., ‘The Industrial Workers of the World’ (a specific union) or ‘St Polycarp’ (a specific martyr). We identified 13,561 additional individuals by checking whether Wiktionary only lists a lemma only as a proper noun and by manually inspecting synsets where all lemmas are capitalized.

V. EVALUATION

Like many efforts in knowledge acquisition and reasoning, the creation of lexical semantic axioms is motivated by a variety of applications, but it is not easily evaluated through them. Instead, it is traditional to rely on human judgements – often those of the authors – to determine the accuracy and appropriateness of the results (as in [27]–[30]).

While it is natural for us to judge a random sample of the resulting axioms, this would not accurately reflect their value for tasks requiring commonsense reasoning. Due to WordNet’s broad lexical coverage, most of the words it includes are rare, including, for instance, specialized scientific and medical terminology. Therefore, as our evaluation set, we took the axioms where both predicates are from a set of “core” synsets.

Table II
DISTRIBUTION OF RATINGS

Rating	Schemata Groups					All	Baseline
	G1	G2	G3	G4	G5		
<i>Best</i> 1	93	85	103	67	51	399	211
2	11	18	8	19	21	77	54
3	6	9	6	12	15	48	46
4	7	5	1	12	20	45	131
<i>Worst</i> 5	3	3	2	10	13	31	158
Average	1.47	1.53	1.21	1.99	2.36	1.72	2.95

These are the union of two standard lists: Boyd-Graber et al.’s “core” WordNet [31] and Izquierdo et al.’s Base Concepts [32]. To get a balanced sample of different relations, we randomly selected 40 axioms for each of the five schemata groups presented in Section IV. For rating, these axioms were shuffled with a baseline interpretation of the same hyponym–hypernym pairs, corresponding to Schema 1, the most common. A random selection of the English verbalizations of our output and baseline output from the evaluation set is presented in Figure 1 and Figure 2.

Judges were asked to evaluate each axiom’s English verbalization based on whether it is a reasonable claim with respect to the general word senses indicated by the definition and examples for the synset. They were instructed that whenever an axiom says ‘amount of’ they should apply the cumulativity test described in the introduction: Does ‘some x and some more x ’ constitute ‘an amount of x ’? They were asked to apply the same test to ‘a(n)’ and ‘every’ phrasings; while ‘amount of’ applied to plurals should be tolerated, ‘every’ or ‘a(n)’ applied to mass terms should not.

Each axiom was rated on a scale of 1 (best) to 5 (worst). The authors each rated the full evaluation set of 400 axioms. The 200 axioms of system output had an average rating of 1.78, while the 200 baseline axioms had an average rating of 3.29. A Pearson correlation of .77 reflects a high level of agreement.

For greater objectivity, it is desirable to also have judges unaffiliated with the work rate the axioms. However, we found it difficult to train judges to be sufficiently sensitive to the property of cumulativity and to resist type-ifying the claims to allow non-basic readings such as ‘a wine is a liquid’. One judge’s ratings for 200 axioms (100 system output, 100 baseline) were well-correlated with the authors (0.6), giving our output an average rating of 1.71 and the baseline 2.59. A second judge’s ratings were less well-correlated (0.45), indicating difficulty in understanding the criteria or in assessing them, but still rated our system’s output better on average (1.47) than the baseline (1.97).

The distribution of ratings for all three judges (200 sys. ratings by each of the authors, 100 by each of the other judges; likewise for the baseline) can be seen in Table II, including a breakdown of the ratings by the axiom schemata groups, showing their relative reliability.

Every amount of reparation is an amount of compensation.

($\forall x$: [x reparation1.n] [x compensation1.n])
— *reparation#1*: compensation (given or received) for an insult or injury
— *compensation#1*: something (such as money) given or received as payment or reparation (as for a service or loss or injury)

Curiosity is a cognitive state.

[(k curiosity1.n) cognitive_state1.n]
— *curiosity#1*: a state in which you want to learn more about something
— *cognitive_state#1*: the state of a person’s cognitive processes

Every cathedral is a church building.

($\forall x$: [x cathedral1.n] [x church_building1.n])
— *cathedral#1*: any large and important church
— *church_building#1*: a place for public (especially Christian) worship

Every abandonment is a rejection.

($\forall x$: [x (plur abandonment1-c.n)] [x (plur rejection1-c.n)])
— *abandonment#1*: the act of giving something up
— *rejection#1*: the act of rejecting something

Every assembly is a gathering.

($\forall x$: [x assembly4.n] [x gathering1.n])
— *assembly#4*: a group of persons who are gathered together for a common purpose
— *gathering#1*: a group of persons together in one place

Every counting is an investigation.

($\forall x$: [x counting1.n] [x (plur investigating1-c.n)])
— *counting#1*: the act of counting; reciting numbers in ascending order
— *investigation#2*: the work of inquiring into something thoroughly and systematically

Computer science is a branch of applied science.

[(k computer_science1.n) branch-of.n (k applied_science1.n)]
— *computer_science#1*: the branch of engineering science that studies (with the aid of computers) computable processes and structures
— *applied_science#1*: the discipline dealing with the art or science of applying scientific knowledge to practical problems

Figure 1. System output from the evaluation set.

VI. REASONING WITH WORDNET AXIOMS

The need for knowledge about entailment relations between entity types has been recognized since the early days of AI (e.g., [33]–[35]). To make commonsense inferences, it is especially important to have the sort of taxonomic knowledge contained in WordNet’s hypernym hierarchy. For example, if we are told ‘Merry is a cat’, a basic reasoning chain is: *Every cat is a feline, every feline is a carnivore, ..., every chordate is an animal*. Therefore, *Merry is an animal*. This process of generalization allows us to apply world knowledge known at a higher level of generality. For instance, if we know *Every animal needs food to live*, we can conclude *Merry needs food to live*, and an intelligent agent might, accordingly, form the goal of feeding her.

A slightly more complex line of reasoning demonstrates the inferential importance of the semantic distinctions we have explored in this paper:

All gold dust is gold.

($\forall x$: [x gold_dust1.n] [x gold3.n])

Gold is a noble metal.

[(k gold3.n) noble_metal1.n]

Every restoration is a repair.

($\forall x$: [x restoration2.n] [x repair1.n])
— *restoration#2*: the act of restoring something or someone to a satisfactory state
— *repair#1*: the act of putting something in working order again

Every physics is a natural science.

($\forall x$: [x physics1.n] [x natural_science1.n])
— *physics#1*: the science of matter and energy and their interactions
— *natural_science#1*: the sciences involved in the study of the physical world and its phenomena

Every pretending is a dissimulation.

($\forall x$: [x pretending1.n] [x dissimulation1.n])
— *pretending#1*: the act of giving a false appearance
— *dissimulation#1*: the act of deceiving

Every sameness is a quality.

($\forall x$: [x sameness1.n] [x quality1.n])
— *sameness#1*: the quality of being alike
— *quality#1*: an essential and distinguishing attribute of something or someone

Every solid is a matter.

($\forall x$: [x solid1.n] [x matter3.n])
— *solid#1*: matter that is solid at room temperature and pressure
— *matter#3*: that which has mass and occupies space

Every encroachment is an influence.

($\forall x$: [x encroachment3.n] [x influence2.n])
— *encroachment#3*: influencing strongly
— *influence#2*: causing something without any direct or apparent effort

Every sand is a dirt.

($\forall x$: [x sand1.n] [x dirt1.n])
— *sand#1*: a loose material consisting of grains of rock or coral
— *dirt#1*: the part of the earth’s surface consisting of humus and disintegrated rock

Figure 2. Baseline output from the evaluation set.

A meta-axiom over mass predicates gives the logical equivalence of our ‘amount of’ verbalizations:

All p is an amount of the kind p (for mass predicate p).

($\forall_{\text{pred } p}$: [$\forall p$ mass-pred]
(all x [[x p] \Leftrightarrow [x amount-of (k p)]]))

And, from our annotation of WordNet, we know

Gold_dust1.n and gold1.n are mass predicates.
[‘gold_dust1.n mass-pred], [‘gold3.n mass-pred]

Therefore,

Every amount of gold dust is an amount of a (certain) noble metal.

($\exists y$: [y noble_metal1.n]
($\forall x$: [x amount-of (k gold_dust1.n)]
[x amount-of y]))

(Rather than *All gold dust is noble metal* or *Every gold dust is a noble metal!*) So if, ignoring tense, we learn

John found some gold dust.

($\exists x$: [x gold_dust1.n] [John.name find.v x])

We can conclude

John found some amount of a noble metal.

($\exists y$: [y noble_metal1.n]
($\exists x$: [x amount-of y]
[John.name find.v x]))

VII. CONCLUSIONS

We have seen that WordNet’s hypernym hierarchy represents a variety of semantically distinct relations. To create lexical axioms suitable for use in a general reasoner, we must identify and formalize these relations. In this paper, we’ve shown that we can use the mass–count distinction to obtain a large number of such axioms, which are judged significantly better than a subsumptive count-noun baseline. With this work, we are releasing an initial collection of lexical axioms² for public use and to establish a baseline for future formalization.

VIII. DISCUSSION AND FUTURE WORK

Our results show that we can significantly improve the reliability of hierarchy axioms extracted from WordNet by attending to the mass–count distinctions among word senses (and some other subtle properties). But our research undertaking has also revealed some systematic difficulties in making logical sense of WordNet hierarchy relations, and these point to interesting possibilities for future work.

Consider this hyponym–hypernym pair:

watching#1: the act of observing; taking a patient look

looking_at#1: the act of directing the eyes toward something and perceiving it visually

Both glosses characterize the word senses in terms of acts, and since ‘act’ is a count noun (both syntactically and conceptually – there can be single acts or multiple, distinct acts), one would expect favorable judgements for an axiom expressing *Every watching is a looking-at*. However, the oddity of the paraphrase *All watchings are lookings-at* makes the count readings of these nouns rather suspect. In fact, natural occurrences like ‘his watching you while you sleep’ (a gerund) or ‘his watching of CNN’ (a deverbal noun) suggest that watching is basically an activity rather than an act. However, this is an elusive intuition as we can easily conceive of bounded *episodes* of any activity, which have the character of acts. Indeed, we can assume that there is a class of “countifying” operators that map activity/process predicates to action/event predicates; for example, adverbials such as ‘for three hours’ accomplish such a transformation (see [36], which systematically treats the semantics of durative and many other types of adverbials; the analysis is for verb phrase adjuncts, but many of the observations carry over to deverbal nouns.)

The relationship between the above two word senses is further obscured by the fact that the synset for ‘watching#1’ also contains the word sense ‘observation#2’, for which the definition as an act seems to fit better: We can naturally speak in the plural of ‘Penn’s observations of the nightly newscasts from Vietnam’, and since the newscasts are bounded events, so are the observations.

This is just one example of the type-shifting (countability-shifting, and, for deverbal nouns, aspectual-category-shifting) transformations that many nouns are susceptible to, and

how this shifting potential can confound intuitive judgements of the validity of hierarchy axioms derived from WordNet synsets. Here are some more shifting operations:

- 1 *Iteration*: e.g., the event predicate ‘sneeze’ becomes a predicate true of the activity of ‘sneezing’ through iteration; WordNet places ‘sneeze#1’ and ‘sneezing#1’ in the same synset.
- 2 *Raising to a kind-level predicate*: e.g., the predicate, ‘wine’, true of quantities of stuff becomes a kind of wine (true, for example, of Merlot or Riesling); WordNet does not have an entry for the latter sense; on the other hand, WordNet’s entry ‘medicine#2’ is grouped into the same synset with ‘drug#1’ – clearly a *kind* of medicine;
- 3 *Conventional portions*: ‘a wine’ or ‘a beer’ can refer to a serving of either stuff, perhaps derivable from the basic predicate by a ‘conventional-portion-of’ or ‘serving-of’ operator. WordNet does not distinguish these senses, but its second entry for ‘tissue’, the synset {tissue#2, tissue_paper#1}, is accompanied by a gloss that is compatible with either a predicate true of any amount of tissue paper, of certain standard portions (‘Please hand me a tissue’), or of certain kinds of paper. Similarly, the gloss for ‘physical exercise’ describes this as activity (thus mass), but the synset also contains ‘workout’, which clearly refers to a conventional bout of exercise.

These observations suggest that future work should look into systematic generation of meaning variants from certain “basic” meanings of nouns. WordNet synsets would then be analyzed (as far as possible automatically) to identify and relate meaning variants within synsets, generated by type-shifting operators of the above types. (The total number of such operators appears to be quite small.) Such a project would fit well with Pustejovsky’s Generative Lexicon project [7], and would enable a more complete and accurate axiomatization of WordNet hierarchy relations. Of course, complete reliability is unattainable, if only because WordNet is not error-free. For example, WordNet relates ‘identity#2’ to hypernym ‘recognition#2’, but the former is defined in terms of individual characteristics and the latter as a process, and there is no way in which characteristics can be construed as a process. But these examples seem to be relatively rare (based on informal observation, perhaps a few out of 100), so there remains considerable scope for extracting relatively reliable, more refined formalized knowledge from WordNet.

ACKNOWLEDGMENTS

This work was supported by NSF grant IIS-0916599 and a subcontract to ONR STTR N00014-10-M-0297.

REFERENCES

- [1] C. Fellbaum, *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- [2] F.J. Pelletier and L.K. Schubert, “Mass expressions,” in *Handbook of Philosophical Logic*, 2nd ed. Dordrecht: Reidel, 2003, vol. 10, pp. 265–350.

² Available at <http://cs.rochester.edu/research/epilog/wn>

- [3] D. Nicolas, "The logic of mass expressions," in *Stanford Encyclopedia of Philosophy*, 2013, <http://plato.stanford.edu/entries/logic-massexpress>.
- [4] D. B. Lenat, "Cyc: A large-scale investment in knowledge infrastructure," *Commun. of the ACM*, vol. 38, pp. 33–8, 1995.
- [5] I. Niles and A. Pease, "Toward a standard upper ontology," in *Proc. Int. Conf. on Formal Ontology in Information Systems (FOIS)*, 2001.
- [6] J. Álvarez, P. Lucio, and G. Rigau, "Adimen-SUMO: Reengineering an ontology for first-order reasoning," *Int. J. on Semantic Web and Information Syst.*, vol. 8, no. 4, 2012.
- [7] J. Pustejovsky, *The Generative Lexicon*. Cambridge, MA: MIT Press, 1995.
- [8] J. Pustejovsky, C. Havasi, J. Littman, A. Rumshisky, and M. Verhagen, "Towards a generative lexical resource: The Brandeis Semantic Ontology," in *Proc. Language Resource and Evaluation Conf. (LREC)*, 2006.
- [9] A. N. Kaplan and L. K. Schubert, "Measuring and improving the quality of world knowledge extracted from WordNet," University of Rochester, Rochester, NY, Tech. Rep., 2001.
- [10] N. Verdezoto and L. Vieu, "Towards semi-automatic methods for improving WordNet," in *Proc. Int. Workshop on Computational Semantics (IWCS)*, 2011.
- [11] N. Guarino and C. Welty, "Ontological analysis of taxonomic relationships," in *Proc. 19th Int. Conf. on Conceptual Modeling*, A. Laender and V. Storey, Eds. Springer-Verlag, 2000.
- [12] A. Gangemi, R. Navigli, and P. Velardi, "The OntoWordNet project: Extension and axiomatization of conceptual relations in WordNet," in *Proc. ODBASE*. Springer-Verlag, 2003, pp. 820–38.
- [13] L. K. Schubert and C. H. Hwang, "Episodic Logic Meets Little Red Riding Hood: A comprehensive, natural representation for language understanding," in *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*, L. Iwanska and S. C. Shapiro, Eds. MIT/AAAI Press, 2000.
- [14] F. Morbini and L. K. Schubert, "Metareasoning as an integral part of commonsense and autocognitive reasoning," in *AAAI Workshop on Metareasoning*, 2008, pp. 155–62.
- [15] L. K. Schubert, B. Van Durme, and M. Bazrafshan, "Entailment inference in a natural logic-like general reasoner," in *AAAI Fall Symp. on Commonsense Knowledge (CSK)*, 2010.
- [16] F. Morbini and L. K. Schubert, "Evaluation of EPILOG: A reasoner for Episodic Logic," in *Proc. Int. Symp. on Logical Formalizations of Commonsense Reasoning*, Jun. 1–3 2009, pp. 103–8.
- [17] B. MacCartney and C. D. Manning, "Modeling semantic containment and exclusion in natural language inference," in *Proc. Int. Conf. on Computational Linguistics (COLING)*, 2008.
- [18] G. Carlson, "A unified analysis of the English bare plural," *Linguistics and Philosophy*, vol. 1, pp. 413–56, 1977.
- [19] F. Bond and C. Vatikiotis-Bateson, "Using an ontology to determine English countability," in *Proc. Int. Conf. on Computational Linguistics (COLING)*, 2002, pp. 1–7.
- [20] T. Baldwin and F. Bond, "A plethora of methods for learning English countability," in *Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, 2003, pp. 73–80.
- [21] T. O'Hara, N. Salay, M. Witbrock, D. Schneider, B. Aldag, S. Bertolo, K. Panton, F. Lehmann, M. Smith, D. Baxter, J. Curtis, and P. Wagner, "Inducing criteria for mass noun lexical mappings using the Cyc KB, and its extension to WordNet," in *Proc. Int. Workshop on Computational Semantics (IWCS)*, Apr. 29–May 4 2003, pp. 34–9.
- [22] M. Lapata and F. Keller, "The Web as a baseline: Evaluating the performance of unsupervised web-based models for a range of NLP tasks," in *Proc. Human Language Technologies – North American Chapter of the Assoc. for Computational Linguistics (HLT-NAACL)*, 2004, pp. 121–8.
- [23] ———, "Web-based models for natural language processing," *ACM Transactions on Speech and Language Processing*, vol. 2, no. 1, pp. 1–31, 2005.
- [24] J. Álvarez, J. Atserias, J. Carrera, S. Climent, E. Laparra, A. Oliver, and G. Rigau, "Complete and consistent annotation of WordNet using the Top Concept Ontology," in *Proc. Language Resource and Evaluation Conf. (LREC)*, 2008, pp. 1529–34.
- [25] H. C. Bunt, *Mass Terms and Model-Theoretic Semantics*, ser. Cambridge Studies in Linguistics Series. Cambridge University Press, 1985.
- [26] T. Brants and A. Franz, "Web 1T 5-gram Version 1," 2006, Linguistic Data Consortium, Philadelphia.
- [27] N. S. Friedland and P. G. Allen, "The Halo pilot: Towards a digital Aristotle," http://projecthalo.com/content/docs/halopilot_vulcan_finalreport.pdf, Vulcan, Inc., Tech. Rep., 2003.
- [28] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, "Open information extraction from the Web," in *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 2007.
- [29] B. Van Durme and L. K. Schubert, "Open knowledge extraction using compositional language processing," in *Proc. Symp. on Semantics in Systems for Text Processing (STEP)*, 2008.
- [30] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka, Jr, and T. M. Mitchell, "Toward an architecture for never-ending language learning," in *Proc. Conf. on Artificial Intelligence (AAAI)*, 2010.
- [31] J. Boyd-Graber, C. Fellbaum, D. Osherson, and R. Schapire, "Adding dense, weighted connections to WordNet," in *Proc. 3rd Global WordNet Meeting*, 2006.
- [32] R. Izquierdo, A. Suárez, and G. Rigau, "Exploring the automatic selection of basic level concepts," in *Proc. Int. Conf. on Recent Advances in Natural Language Processing (RANLP)*, 2007.
- [33] R. A. Amsler and J. S. White, "Development of a computational methodology for deriving natural language semantic structures via analysis of machine-readable dictionaries," National Science Foundation, Tech. Rep. MCS77-01315, 1979.
- [34] M. S. Chodorow, R. J. Byrd, and G. E. Heidorn, "Extracting semantic hierarchies from a large on-line dictionary," in *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, Jul. 8–12 1985, pp. 299–304.
- [35] R. Wilensky, "Knowledge acquisition and natural language processing," in *Foundations of Knowledge Acquisition: Cognitive Models of Complex Learning*, A. Meyrowitz, Ed. Boston, MA: Kluwer, 1993.
- [36] C. H. Hwang and L. K. Schubert, "Interpreting tense, aspect, and time adverbials: a compositional, unified approach," in *Proc. Int. Conf. on Temporal Logic*, D. M. Gabbay and H. J. Ohlbach, Eds. Springer-Verlag, 1994, pp. 238–64.