Nominal Compound Interpretation by Intelligent Agents

Marjorie McShane
Stephen Beale
Petr Babkin

Nominal Compound Interpretation by Intelligent Agents

Marjorie McShane, *Rensselaer Polytechnic Institute,* Stephen Beale, *Rensselaer Polytechnic Institute,* Petr Babkin, *Rensselaer Polytechnic Institute*

Abstract

This paper presents a cognitively-inspired algorithm for the semantic analysis of nominal compounds by intelligent agents. The agents, modeled within the OntoAgent environment, are tasked to compute a full context-sensitive semantic interpretation of each compound using a battery of engines that rely on a high-quality computational lexicon and ontology. Rather than being treated as an isolated "task", as in many NLP approaches, nominal compound analysis in OntoAgent represents a minimal extension to the core process of semantic analysis. We hypothesize that seeking similarities across language analysis tasks reflects the spirit of how people approach language interpretation, and that this approach will make feasible the long-term development of truly sophisticated, human-like intelligent agents. The initial evaluation of our approach to nominal interpretation is promising, and suggests one feature of nominal compounds that has been long-recognized by linguists but runs counter to much recent work on machine-learning-oriented approaches to NN analysis: many nominal compounds are fixed expressions, requiring individual semantic specification at the lexical level.

1 Introduction

A nominal compound (NN) is a sequence of two or more nouns in which one is the head and the other(s) are modifiers: e.g., *glass bank*. One of the central challenges in automatically interpreting compounds is that the nouns can be polysemous, making the task not only to determine the semantic relationship between them, but also the contextually appropriate meaning of each noun. For example, although *glass bank* might be readily recognized, outside of context, as ambiguous by human readers (a coin storage unit made of glass; a slope made of glass; a storage unit for glass; a financial institution with a prominent architectural feature made of glass; etc.), the compounds *pilot program* and *home life* might seem unambiguous. However, there are other readings – which are surprisingly plausible when one thinks about it – that are equally available to language processing systems: *pilot program* could mean a program for the benefit of airplane pilots and *home life* could refer to the length of time that a dwelling is suitable to be lived in (by analogy with *battery life*). So lexical disambiguation is as central to nominal compound analysis as is the establishment of the relationship between the nouns.

Nominal compounding has been pursued by descriptive linguists, psycholinguists, and practitioners of natural language processing (NLP). By way of introduction, we will first provide a brief, interpretive glimpse into some lines of work pursued by each of these communities, without repeating the fine surveys already available in, e.g., Lapata 2002, Tratz and Hovy 2010, and Lieber and Štekauer 2009.

Descriptive linguists have primarily investigated constraints on the form of NN compounds and the inventory of relations that can hold between the component nouns. They have posited anywhere from 6 to 60 to "innumerable" necessary relations, depending on their evaluation of an appropriate grain-size of semantic analysis. They do not pursue algorithms for disambiguating the component nouns, presumably because the primary consumers of linguistic descriptions are people who carry out such disambiguation automatically; however, they do pay well-deserved attention to the fact that NN interpretation requires a discourse context, as illustrated by Downing's (Downing 1977) now classic "apple-juice seat" example. Psycholinguists, for their part, have found that the speed of NN processing increases if one of the component nouns occurs in the immediately preceding context (cf. Gagné and Spalding 2006). Speed gains by people in coreference-supported contexts mirror gains in confidence for intelligent agent systems attempting to disambiguate component nouns, a topic to be discussed

further below.

Within recent mainstream NLP, most practically-oriented work on NN compounding belongs to the knowledge-lean paradigm. Practitioners typically select a medium-sized subset of relations of interest and train their systems to automatically choose the relevant relation during analysis of compounds taken outside of context – i.e., presented as a list. Two methods have been used to create the inventory of relations: developer introspection, often with iterative refinement (e.g., Moldovan et al. 2004), and crowd-sourcing, also with iterative refinement (e.g., Tratz and Hovy 2010). A recent direction of development involves using paraphrases as a proxy for semantic analysis: i.e., a paraphrase of a NN that contains a preposition or a verb is treated as the meaning of that NN (e.g., Kim and Nakov 2011). Evaluations of knowledge-lean systems typically compare machine performance with human performance on a relation-selection or paraphrasing task.

In most contributions within the knowledge-lean NLP paradigm, the semantics of the component nominals is not directly addressed: i.e., semantic relations are used to link uninterpreted nominals. Although this might seem incongruous from a linguistic perspective, one can find motivations for pursuing NN compounding thus defined. (1) The developers' purview can be a narrow, technical domain that includes largely monosemous nouns (e.g., medicine, as in Rosario and Hearst 2001), making nominal disambiguation not a central problem.[1] (2) The development effort can be squarely application-oriented, with success being defined as near-term improvement to an end system, with no requirement that all aspects of NN analysis be addressed. (3) The work can be method-driven, meaning that its goal is to improve our understanding of machine learning itself, with the NN dataset being of secondary importance. (4) Systems can be built to participate in a field-wide competition, for which the rules of the game are posited externally (cf. the "Free paraphrases of noun compounds" task of SemEval-2013, Hendrickx et al. 2013). Understanding this broad range of developer goals helps not only to put past work into perspective, but also to explain why the "full semantic analysis" approach described here does not represent an evolutionary extension to what came before; instead, it addresses a different problem space altogether. It is closest in spirit to the work of Moldovan et al. 2004, who also undertake nominal disambiguation; but whereas they implement a pipeline – word sense disambiguation followed by relation selection – we combine these aspects of analysis,

---

[1]Similarly, Lapata 2002 developed a probabilistic model covering only those 2-noun compounds in which N1 is the underlying subject or direct object of the event represented by N2: e.g., car lover.

TABLE 1 OntoAgent Analyses of Compounds

No.	Example	Full NN analysis by OntoAgent
1	cooking pot	POT INSTRUMENT-OF COOK
2	eye surgery	PERFORM-SURGERY THEME EYE
3	cat food	FOOD THEME-OF INGEST (AGENT CAT)
4	shrimp boat	BOAT LOCATION-OF CATCH-FISH (THEME SHRIMP)
5	plastic bag	BAG MADE-OF PLASTIC
6	court order	ORDER AGENT LEGAL-COURT
7	gene mutation	MUTATE THEME GENE
8	papilloma growth	CHANGE-EVENT THEME PAPILLOMA (PRECONDITION SIZE ($<$ SIZE.EFFECT))
9	headache onset	HEADACHE PHASE BEGIN
10	pet spray	LIQUID-SPRAY THEME-OF APPLY (BENEFICIARY PET)

TABLE 2 Relation Selection Analyses of Compounds

No.	Example	Relation selection from an inventory
1	cooking pot	perform/engage_in[T]
2	eye surgery	modify/process/change [T]
3	cat food	consumer + consumed [T]
4	shrimp boat	obtain/access/seek]T]
5	plastic bag	substance/material/ingredient + whole [T]
6	court order	communicator of communication [T]
7	gene mutation	defect [R]
8	papilloma growth	change [R]
9	headache onset	beginning of activity [R]
10	pet spray	for [L]

additionally incorporating tasks such as reference-supported sense disambiguation and learning new words.

2 NN Interpretation with and without Nominal Semantics

As a prelude to describing the OntoAgent approach to NN analysis, let us briefly consider best-case NN analysis results from different paradigms. A sampling of examples is shown in Tables 1 and 2. Table 1 shows optimal results of OntoAgent analysis, whereas Table 2 shows examples from three other paradigms: Tratz and Hovy 2010 [T], Rosario and Hearst 2001 [R] and Levi 1979 [L].

A practical aside. The comparisons below suggest that the full analysis sought in OntoAgent is semantically superior to the relation-selection approach pursued elsewhere. This does not mean that it is preferable along every practical axis: it is more expensive, its development is a long-term endeavor that involves many aspects of language and world modeling, and it is more difficult to evaluate using standard metrics. So, we are not suggesting that this is the only rational way to pursue the automation of NN analysis. We are, however, suggesting that if systems can analyze the full meaning of NNs in context, the results will better support human levels of machine reasoning.

With this in mind, consider the following comparisons:

1. The OntoAgent analyses include disambiguation of the component nouns along with identification of the relation between them, whereas relation-selection approaches address only the relation itself.

2. The OntoAgent analyses are written in an unambiguous, ontologically grounded metalanguage (concepts are written in small caps), whereas the relation-selection approaches use potentially ambiguous English words and phrases.

3. In 2, 7, 8 and 9, the meaning of the "relation" in relation-selection approaches is actually not a relation at all but, rather, the meaning of the second noun or its hypernym: e.g., growth is a kind of change. By contrast, since the OntoAgent treatment involves full analysis of all aspects of the compound, the meaning of each of the nouns is more naturally included in the analysis.

4. The relation-selection approach can merge relations into supersets that are not independently motivated: e.g., [T]'s *obtain/access/seek*.[2] In OntoAgent, by contrast, every relation available in the independently developed ontology is available for use in compounding analysis – there is no predetermined list of "compounding relations". This harks back to opinions expressed decades ago that practically any relationship could be expressed by a nominal compound (e.g., Finin 1980).

5. Relation-selection approaches can include unbalanced sets of relations: e.g., *consumer + consumed* [T] has been promoted to the status of a separate relation but many other agent + theme combinations have not.

---

[2][T] prioritized achieving interannotator agreement in a Mechanical Turk experiment, and this methodology influenced their final inventory of relations. Thanks to them for sharing their data.

6. Relation-selection approaches do not support the recognition of paraphrases. By contrast, within OntoAgent, the same meaning representation is generated whether the input is *headache onset, the onset/beginning/start of the headache, someone's headache began/started*, etc. Such recognition of paraphrases is, of course, central to the paraphrasing approaches mentioned earlier, but they do not disambiguate the nominal elements of the original NN or its paraphrase.

In sum, for OntoAgents there is no isolated NN task that exists outside of overall semantic analysis of a text. OntoAgents need to compute the full meaning of compounds along with the full meaning of everything else in the discourse, with the same set of challenges encountered at every turn. For example:

- Processing the elided relations in NN compounds is similar to processing semantically underspecified lexemes. In NNs (e.g., *physician meeting*), the relation holding between the nouns is elided and must be inferred; but in paraphrases that contain a preposition (e.g., *meeting of physicians*), the preposition can be so polysemous that it provides little guidance for interpretation anyway. Both of these formulations require the same reasoning by OntoAgents to arrive at the interpretation: MEETING-EVENT AGENT SET (MEMBER-TYPE PHYSICIAN) (CARDINALITY > 1).

- Unknown words can occur in any aspect of input. Encountering out-of-lexicon words is a constant challenge for agents, and it can be addressed using the same types of machine learning processes in all cases.

- Many compounds are lexically idiosyncratic. Although past work has considered compounds like *drug trial* and *tea cup* to be productive collocations, they arguably are not, instead representing specific elements of a person's world model whose full meaning cannot be arrived at by compositional analysis. We agree with ter Stal and van der Vet 1994 that much more lexical listing is called for in treating compounds. Citing just a short excerpt from their larger discussion:

  "In natural science, [room temperature] means precisely 25 degrees Centigrade. A process able to infer this meaning would have to make deductions involving a concept for room, its more specific interpretation of room in a laboratory, and the subsequent standardisation that has led to the precise meaning given above. All these concepts play no role whatsoever in the rest of the system. That is a high price to pay for the capacity to infer the meaning of [room tem-

perature] from the meanings of room and temperature. Thus, [room temperature] is lexicalized." (ibid)

- Analysis should not introduce semantic ellipsis. The relation-selection method often introduces semantic ellipsis. For example, [T] analyze *tea cup* as *cup* with the *purpose of tea*; but only events can have purposes, so this analysis introduces ellipsis of an event like *drink*. Similarly, *shrimp boat* [T] is cited as an example of *obtain/access/seek*, but the boat is not doing any of this, it is the location of fisherman who are doing this. If intelligent agents are to be armed to reason about the world like people do, then they need to be furnished with non-elliptical analyses, or else configured to subsequently recover the meaning of those ellipses.

In sum, viewing NN compounds within the context of broad-scale semantic analysis is a different task from what has been pursued to date in past descriptive and NLP approaches. Let us turn now to how we are preparing OntoAgents to undertake this task.

3 The OntoAgent Environment

OntoAgent is a cognitive architecture that supports the development of multi-functional, language endowed intelligent agents (see McShane and Nirenburg 2012 and references therein). In OntoAgent, all physiological, general cognitive and language processing capabilities of all intelligent agents rely on the same ontological substrate, the same organization of the fact repository (agent memory of assertions) and the same approach to knowledge representation.

The ontology is organized as a multiple-inheritance hierarchical collection of frames headed by concepts that are named using language-independent labels. It currently contains approximately 9,000 concepts, most of which belong to the general domain. Consider an excerpt from the ontological frame for DRUG-DEALING as we describe salient properties of the ontology.

DRUG-DEALING

	AGENT	default	CRIMINAL, DRUG-CARTEL
		sem	HUMAN
		relaxable-to	SOCIAL-OBJECT
	THEME	default	ILLEGAL-DRUG
	INSTRUMENT	sem	MONEY

Concepts divide up into EVENTs, OBJECTs and PROPERTies. PROP-

ERTies are primitives, which means that their meaning is understood
to be grounded in the real world without the need for further ontolog-
ical decomposition. The facets *default, sem,* and *relaxable-to* allow for
recording more and less typical constraints on property values.

Since the OntoAgent ontology is language independent, its link to
any natural language must be mediated by a lexicon. Consider, for
example, the first two verbal senses for address, shown below.

```
address
  address-v1
  anno
    definition  "to talk to, give a speech to"
    example "He addressed the audience."
  syn-struc
    subject       $var1
    v             $var0
    directobject  $var2
  sem-struc
    SPEECH-ACT
    AGENT        ^$var1   (sem HUMAN)
    BENEFICIARY ^$var2   (sem HUMAN)
                         (relaxable-to ANIMAL)

  address-v2
    anno
     definition  "to consider, think about"
     example  "He addressed the problem promptly."
    syn-struc
      subject       $var1
      v             $var0
      directobject $var2
    sem-struc
      CONSIDER
        AGENT    ^$var1  (sem HUMAN)
        THEME    ^$var2  (sem ABSTRACT-OBJECT)
```

Syntactically, both senses expect a subject and a direct object in the
active voice, filled by the variables $var1 and $var2, respectively. How-
ever, in address-v1, the meaning of the direct object (^$var2) is con-
strained to a HUMAN (or, less commonly, ANIMAL), whereas in address-

v2 the meaning of the direct object is constrained to an ABSTRACT-OBJECT. This difference in constraints permits the analyzer to disambiguate: if the direct object is abstract, as in *He addressed the problem*, then *address* will be analyzed as CONSIDER using address-v2; by contrast, if the direct object is HUMAN, as in *He addressed the audience*, then *address* will be analyzed as SPEECH-ACT using address-v1.

The OntoAgent text analyzer takes as input natural language text and generates disambiguated, ontologically grounded interpretations, called text meaning representations (TMRs), that are well suited to machine reasoning. Basic TMRs include the results of lexical disambiguation and the establishment of the semantic dependency structure, whereas extended TMRs include the results of reference resolution, the interpretation of indirect speech acts, and other discourse-level aspects of language processing. As an example of text processing, consider the TMR for the input *Charlie watched the baseball game.*

VOLUNTARY-VISUAL-EVENT-1

	AGENT	HUMAN-1
	THEME	BASEBALL-GAME-1
	TIME	(before find-anchor-time)
	textstring	"watched"
	from-sense	watch-v1

HUMAN-1

	AGENT-OF	VOLUNTARY-VISUAL-EVENT-1
	HAS-NAME	"Charlie"
	textstring	"Charlie"
	from-sense	*personal-name*

BASEBALL-GAME-1

	THEME-OF	VOLUNTARY-VISUAL-EVENT-1
	textstring	"baseball_game"
	from-sense	baseball_game-n1

Although this level of text analysis requires high-quality, machine-tractable knowledge bases and equally well-crafted processors – all of which, to date, are manually acquired in the OntoAgent environment – once they have been developed, they can support semantically-oriented analysis of all types of language input, include nominal compounds.

4 The NN Analysis Algorithm

When the OntoAgent text processor encounters an NN, it calls an ordered set of functions that offer decreasing levels of confidence in their

output. Currently, the first analysis that achieves a threshold of confidence when incorporated into the larger context is accepted and NN processing stops; however, an alternative control strategy would be to apply all analysis functions to each input and, at the end, select the one with the highest score.

To illustrate what we mean by "decreasing levels of confidence", let us consider the extremes: if a compound is recorded in the lexicon as a head entry (e.g., attorney_general) and its meaning aligns with the selectional constraints imposed by its selecting head (as in *The attorney general announced the meeting* (the SPEECH-ACT instantiated by *announce* expects a HUMAN AGENT)), that meaning is incorporated into the text meaning representation with the highest degree of confidence. By contrast, if one or more nouns in the compound is an unknown word, then the system can attempt machine learning of the word, with necessarily lower confidence in the resulting overall analysis. Between these extremes are many levels of analysis that centrally rely on the lexical and ontological knowledge bases of OntoAgent.[3]

Pseudocode for the NN analysis algorithm is presented below, with line numbers used for reference in the discussion to follow. This algorithm covers the full inventory of foreseeable eventualities, a top-down development strategy that we consider a cornerstone of building systems that have a naturally long trajectory of development. The algorithm is being implemented in stepwise fashion along with ever improving engines for supporting functions such as lexical disambiguation, reference-supported sense disambiguation, and learning new words.[4]

1. Detect nominal compound N1N2.
2. If the function *find-contextually-appropriate-NN-head-entry* returns NN-INTERP
3. then use NN-INTERP.
4. Else if the lexicon contains at least one sense of N1 and N2
5. then create a set of all senses (i.e., ontological interpretations) of N1 and of N2.
6. Use reference-supported sense disambiguation for initial scoring of the likelihood of each available interpretation.
7. Pass on all available senses, with their reference-oriented scores, for subsequent evaluation. [*First, narrowly specified patterns are tested.*]

---

[3]Of course, OntoAgent is not the first NLP approach to incorporate handwritten rules; see Vanderwende 1994 for discussion.

[4]The format of the page precludes the traditional indentation-oriented formatting convention for algorithms. In the presentation that follows, explanatory roadmarks in bracketed italics are provided for guidance at select points.

8. If the function *find-contextually-appropriate-lexically-anchored-phrasal* returns NN-INTERP then use NN-INTERP.

9. Else if the function *find-contextually-appropriate-NN-pattern* returns NN-INTERP then use NN-INTERP.

10. Else if the function *find-PP-shift-to-NN* returns NN-INTERP then use NN-INTERP. [*This ends testing against narrowly specified patterns.*]

11. Else if the function *find-contextually-appropriate-free-sense-combination* returns NN-INTERP then use NN-INTERP. [*If there is no confident analysis of the NN together, the system turns its attention to N2 by itself, the head of the compound, whose interpretation must fit into the clause-level text meaning.*]

12. Else if the function *find-contextually-appropriate-N2-interpretation* returns N2-INTERPR then

13. if reference-resolution suggests N1-INTERP then return N1-INTERP RELATION N2-INTERP.

14. Else if the function *launch-ML-for-N1* suggests N1-INTERP then return N1-INTERP RELATION N2-INTERP.

15. Else, either accept residual ambiguity of N1 (return [set of all senses of N1] RELATION N2-INTERP) or use application-specific recovery strategy (e.g., ask human collaborator for clarification). [*This ends processing of NNs in which both N1 and N2 are in the lexicon and N2 can be confidently disambiguated. Now the system treats cases in which N2 is known but N1 is unknown.*]

16. Else if the lexicon contains senses of N2 [*but not N1*] then

17. if the function *find-contextually-appropriate-N2* returns N2-INTERP then

18. if ML is **unavailable**, then resolve meaning as [uninterpreted N1] RELATION N2-INTERP.

19. Else [*ML is available*] if the function *find-N2-informed-ML-of-compound* returns NN-INTERP then use NN-INTERP.

20. Else either accept residual ambiguity (i.e., return [set of all senses of N1] RELATION N2-INTERP) or use application-specific recovery strategy (e.g., ask human collaborator for clarification).

21. Else [*the agent cannot confidently disambiguate N2*] if the function *find-raw-ML-of-compound* returns NN-INTERP, then use NN-INTERP.

22. Else either return [uninterpreted N1] RELATION [set of analyses of N2] or use application-specific recovery strategy.

23. Else [*the lexicon does not contain N2*] if N1 is in the lexicon and has a confident reference-informed disambiguation then

24. if the function *launch-N1-informed-ML-of-compound* returns NN-INTERP then use NN-INTERP.

25. Else [*N1 won't help enough or is also absent from the lexicon*] if the function *launch-raw-ML-of-compound* returns NN-INTERP then use NN-INTERP.

26. Else use application-specific recovery strategy.

Line 1. Nominal compounds are detected during syntactic parsing which, in OntoAgent, is carried out by the Stanford dependency parser (de Marneffe et al. 2006). False positives are not uncommon, a fact we mention only to contrast the challenges of end-to-end language analysis (in which upstream errors regularly confound downstream processing) with the simplifications of competition-oriented task specifications. We will refer to the elements of the compound as N1 and N2.

Line 2. The system determines whether the compound is recorded as a head entry in the lexicon, like attorney_general-n1 (i.e., the first nominal sense of *attorney general*). If it is, the recorded semantic interpretation – here, the ontological concept ATTORNEY-GENERAL – is evaluated in context. For example, if the input is *The attorney general announced the meeting*, the analyzer will consider all lexical senses of *announce* and all senses of *meeting* and combine them with ATTORNEY-GENERAL to create the best meaning representation. That process is carried out by an approach to dynamic programming called Hunter Gatherer (Beale 1997). In our example, the optimal analysis is represented in the following TMR:

SPEECH-ACT-1

	AGENT	ATTORNEY-GENERAL-1
	THEME	MEETING-EVENT-1
	TIME	($<$ find-anchor-time)

ATTORNEY-GENERAL is accepted as the analysis of attorney general because it perfectly fulfills the ontologically-recorded expectation that the AGENT of a SPEECH-ACT is HUMAN, ATTORNEY-GENERAL being an ontological descendant of HUMAN.

Lines 2-3. Throughout the algorithm, all candidate analyses are evaluated within the context of the larger input. If the automatically generated contextual score is above a set threshold of quality, then the function outputs NN-INTERP, which indicates a **successful** interpretation. If the contextual score is not above a set threshold of quality, the function does not output an interpretation at all and the next function in the series is called. As mentioned earlier, this control structure

could be modified to return all possible analyses with their respective scores.

Line 4. If the compound is not lexically recorded as a headword, or if it is but the lexically recorded meaning does not fit the context, then the next step depends upon whether or not the component nouns are in the system's lexicon. We will start with the case in which both nouns are known, but with the following caveat: it is always possible that a given word is known (i.e., is in the lexicon) but that the meaning relevant for the compound is missing. OntoAgent addresses this realistic, open-world assumption directly, by evaluating candidate analyses in context; however, it would be unwise to underestimate the potential for related errors. For example, if an American-oriented lexicon contained only one meaning of *boot* (footwear) but a British text included the phrase *I hope there's no boot damage* to mean "damage to the trunk of the car", then it would be difficult for the agent to determine that the its recorded meaning "footwear" was not intended – after all, both footwear and car trunks are subject to damage, and people are free to worry about either one. For practical reasons, OntoAgents do not assume that every encountered lexeme could have additional unrecorded meanings that must be learned.

Line 5. If both words are recorded in the lexicon, then the agent builds a list of all of their senses – i.e., the concept-based descriptions recorded in the sem-strucs of their entries. It is the word senses, realized as concepts, not the English strings, that will be evaluated throughout.

Line 6. Next, reference resolution procedures are applied to each candidate sense, using the OntoAgent approach to reference resolution described in McShane 2009 and McShane and Nirenburg 2013.[5] The agent attempts to determine whether the preceding discourse contains a sponsor for either of the compound's nouns, which can serve as a strong guide for lexical disambiguation. The analysis involves a battery of heuristics that measure such features as (a) the semantic distance between the meanings of all candidate sponsors in the window of coref-

---

[5]According to our theory of reference, which has little in common with the "co-reference resolution task" of mainstream NLP, all referential nouns and verbs are subject to reference resolution procedures. We define the resolution of reference as the linking of overt (or elided) mentions of objects and events to anchors in an agent's memory of object and event instances; we view textual coreference resolution as being a possible but not necessary intermediate step in that process. The sponsor for a referring expression need not be coreferential with it. For example, in the sentence *She went to the mall and stopped by the jewelry store first*, *store* is, ontologically, in a meronymic relationship with *mall*. This so-called "bridging" relationship helps to contextually disambiguate *store*, which, in this context, means retail establishment rather than quantity/stash.

erence and all candidate meanings of the noun in question and (b) the text distance between the candidate sponsors and the noun. Consider the analysis of *dog beds* in the input *Next week I'm adopting a puppy! I wonder how many dog beds I should buy for him?* When analyzing *dog beds*, the agent has 3 candidate senses for dog – DOG (a canine), HU-MAN (style: derogatory), and FOLLOW. It also has two candidate senses for bed – BED-FOR-SLEEPING and FLOWER-BED. The availability of the candidate sponsor *puppy*, which is analyzed as DOG (AGE (LESS-THAN 1 YEAR)), is a strong reference-oriented vote for the DOG interpretation of *dog* in *dog beds*. By contrast, if the 2nd sentence of that example were discourse-initial, there would be no reference-oriented reason to prefer the "canine" meaning over the "(derogatory) human" reading.

Positing disambiguation-supporting candidate reference relations includes assigning a level of confidence to each hypothesis. This score will be combined with the confidence score associated with candidate analyses to ultimately select the best overall NN analysis. The numerical aspects of scoring procedures are not trivial: e.g., how does use of a high-confidence NN analysis rule without reference-supported nominal disambiguation compare with a low-confidence NN rule with reference-supported nominal disambiguation? Our initial approach to scoring is coarse-grained: f all heuristic evidence strongly points to the given analysis, then the agent accepts it; if all evidence disfavors the given analysis, the agent rejects it; if the evidence is mixed, then the agent detects its own lack of certainty and seeks a higher confidence analysis.

Line 7. The reference-informed scores of each candidate analysis of N1 and N2 are recorded as input to subsequent processing.

Line 8. The next most confident NN analysis strategy involves detecting if the lexicon includes a sense of N2 that syntactically permits a compounding structure (i.e., has an optional N1 slot), and that lexically or semantically constrains the meaning of N1 so that the relationship between them is predictable. This class includes two subclasses of phenomena.

Subclass 1: The syn-struc of the entry for N2 indicates that a NN structure is possible, and explicitly indicates which *word(s)* can occupy N1. We have relatively few of these at the moment, covering such compounds as *I myself, you yourself*, etc.; and *degree Celsius, degree centigrade*, etc. The reason these are not recorded as multi-part head words is that, in our approach, the first element of a multi-word headword cannot inflect or be separated by punctuation, whereas here we'd like to permit *I, myself,* (with commas) and *degrees Celsius* (the plural).

Subclass 2: The syn-struc of the entry for N2 permits a compounding structure, and the sem-struc indicates semantic constraints on N1. This

class is best explained through examples.

· In compounds of the pattern *X fishing*, if X is a kind of FISH, then the compositional meaning is FISHING-EVENT THEME X: e.g., *trout fishing* means FISHING-EVENT THEME TROUT.
· In compounds of the pattern *Y detective*, if Y is a kind of ILLEGAL-DRUG, WEAPON or CRIMINAL-ACTIVITY, then the compositional meaning is DETECTIVE AGENT-OF INVESTIGATE (THEME Y): e.g., *homicide detective* means DETECTIVE AGENT-OF INVESTIGATE (THEME MURDER).
· In compounds of the pattern *Z hospitalization*, if Z is a kind of PATHOLOGIC-FUNCTION, then the compositional meaning is HOSPITAL-STAY CAUSED-BY Z: e.g., *appendicitis hospitalization* means HOSPITAL-STAY CAUSED-BY APPENDICITIS.

Figure 1 shows a screen shot of the lexical entry for *detective* that prepares the system to analyze compounding configurations using this word. The syn-struc says that the head ($var0) is obligatory but can optionally be used with a preceding nominal ($var1) in a compounding structure. The sem-struc says that head means DETECTIVE; additionally it indicates that if the input does include $var1, and if some meaning of $var1 (remember, it can be polysemous) fits the listed ontological constraints ILLEGAL-DRUG, WEAPON, CRIMINAL-ACTIVITY, then the interpretation of the whole compound includes the elided event INVESTIGATE, as described above and shown in the lower right of the screen shot. If, by contrast, the input were *university detective*, then "university" would not fit these ontological constraints for N1 and the compound would not be analyzed using this lexically recorded pattern; instead, the system would continue through the NN analysis algorithm in search of another analysis.

The OntoAgent lexicon currently includes dozens of such lexically-anchored patterns and, over time, should be expanded to include hundreds if not thousands, since this expectation-oriented approach to knowledge engineering offers high precision analyses for what it covers.

Line 9. At this stage, the analyzer evaluates whether pairs of the available nominal senses correspond to recorded ontological patterns, like TEMPORAL-UNIT + EVENT (e.g., night flight). From the point of view of the analyzer, this process is very similar to the one in Line 8, and renders analyses of a similar level of confidence. The main difference involves the types of knowledge structures consulted and how they are acquired. Whereas lexically-anchored patterns can be recorded in the lexicon, ontological patterns cannot since there is no lexeme to
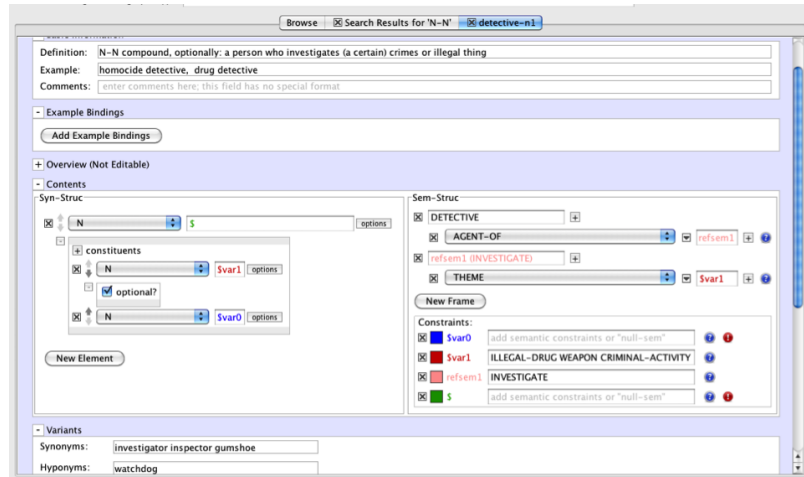
FIGURE 1 Lexical acquisition interface.

serve as the head entry. Instead, a dedicated repository of ontological patterns must be developed. Examples can be functionally divided into those showing "unconnected constraints" and those showing "connected constraints". The latter indicates that the candidate meanings of one of the nouns are tested as a property filler against the ontological expectations of the candidate meanings of the other noun.

Examples of Unconnected Constraints

a. If N1 is TEMPORAL-UNIT and N2 is EVENT then N2 TIME N1: Tuesday flight > FLY-EVENT TIME TUESDAY.

b. If N1 is EVENT and is N2 is EVENT then N2 THEME N1: dream analysis > ANALYZE THEME DREAM.

c. If N1 is ANIMAL-DISEASE or ANIMAL-SYMPTOM and N2 is HUMAN (not MEDICAL-ROLE) then N2 EXPERIENCER-OF N1: polio sufferer > HUMAN EXPERIENCER-OF POLIO.

d. If N1 is SOCIAL-ROLE and N2 is SOCIAL-ROLE then the compound means HUMAN HAS-SOCIAL-ROLE N1 & N2): physician neighbor > HUMAN HAS-SOCIAL-ROLE PHYSICIAN & NEIGHBOR.

e. If N1 is FOODSTUFF and N2 is PREPARED-FOOD then N2 CONTAINS N1: papaya salad > SALAD CONTAINS PAYAYA-FRUIT.

Examples of Connected Constraints

f. If N1 is EVENT and N2 is ANIMAL, and N2 is a "default" or "sem" AGENT of N1, then N1 AGENT N2: cleaning lady > CLEAN-EVENT AGENT HUMAN (GENDER FEMALE).

g. If N1 is EVENT and N2 is an ontologically recorded "sem" or "default" INSTRUMENT of N1 then N1 INSTRUMENT N2: cooking pot > COOK INSTRUMENT POT-FOR-FOOD.

h. If N1 is OBJECT and N2 is a filler of the HAS-OBJECT-AS-PART slot of N1 then N2 PART-OF-OBJECT N1: oven door > DOOR PART-OF-OBJECT OVEN.

i. If N1 is EVENT and is N2 is EVENT, and N2 is a filler of the HAS-EVENT-AS-PART slot of N1, then N2 PART-OF-EVENT N1: ballet intermission > INTERMISSION PART-OF-EVENT BALLET.

j. If N2 is EVENT and N1 is a "default" or "sem" THEME of N2 then N2 THEME N1: photo exhibition > EXHIBIT THEME PHOTOGRAPH.

k. If N2 is described in the lexicon as HUMAN AGENT-OF EVENT-X (e.g., "teacher" is HUMAN AGENT-OF TEACH), and N1 is a "default" or "sem" THEME of X (e.g., PHYSICS is a "sem" filler for the THEME of TEACH), then the NN analysis is HUMAN AGENT-OF X (THEME N1): e.g., physics teacher > HUMAN AGENT-OF TEACH (THEME PHYSICS); home inspector > HUMAN AGENT-OF INSPECT (THEME PRIVATE-RESIDENCE); stock holder > HUMAN AGENT-OF OWN (THEME STOCK-FINANCIAL).

l. If N1 is PHYSICAL-OBJECT and N2 is PHYSICAL-OBJECT and N1 is a "default" or "sem" filler of the MADE-OF slot of N2 then N2 MADE-OF N1: denim skirt > SKIRT MADE-OF DENIM.

m. If N2 is PROPERTY and N1 is a legal filler of the DOMAIN of N2 then N2 DOMAIN N1: ceiling height > HEIGHT DOMAIN CEILING.

These patterns not only offer high-confidence analyses of the relation binding the nominals, they also disambiguate the nominals. For example, although *papaya* can mean PAPAYA-FRUIT or PAPAYA-TREE, in *papaya salad* it can be disambiguated to PAPAYA-FRUIT in order to match pattern (e). It is important to emphasize that these rules seek only high-confidence ontological relations, defined using the "default" and "sem" facets of the ontology. If a compound is semantically idiosyncratic enough that it would correspond only to the "relaxable-to" facet of recorded ontological constraints, then it is not handled at this point in analysis. For example, although TEACH expects its THEME to be an academic subject or skill, the NN *hooliganism teacher* does have a meaning – a person who teaches others how to be hooligans. The

ontology allows for this by including in the frame for TEACH that its theme is "relaxable-to" any EVENT; but a corresponding analysis will not be hypothesized at this stage due to its low confidence.

One might ask why we ever record lexically-anchored patterns (Line 8) since all such patterns could be generalized to ontological concepts: e.g., rather than recording "FISH fishing" under the head word fishing-n1, we could record the pattern FISH FISH-EVENT in the ontological pattern repository. In the latter scenario, when the system encountered the word *fishing*, it would recognize it as FISH-EVENT, resulting in the same analysis. The reason for the split has primarily to do with convenience for acquirers. If a given word, like *fishing*, is often used in compounds, and if it has no synonyms or only a few synonyms that can readily be listed in its "synonyms" field, then it is cognitively simpler and faster for the acquirer to record the information in the lexicon under *fishing*, rather than switch to the compounding repository and seek concept-level generalizations. This is just one example of the larger issue of how to divvy up meaning description across lexicon and ontology, a topic discussed in McShane et al. 2005.

Line 10. The next function again involves lexical search, this time attempting to determine if the NN could be a paraphrase of a N+PP construction that is recorded in the lexicon: e.g., *restaurant chain* can be readily analyzed if *chain of X* is recorded in the lexicon (and, of course, if *restaurant* fits the listed semantic constraints) (cf. Isabelle 1984). Recording the meanings of typical N+PP collocations is done as a matter of course in OntoAgent to assist the analyzer with the tremendously difficult challenge of disambiguating prepositions. Continuing with the example *chain of X*, the lexical sense chain-n2 explicitly lists the PP adjunct "of X" and indicates that if it is present – and if X is of an appropriate semantic type – then the whole structure is to be interpreted as SET MEMBER-TYPE X. It would be extremely difficult for the semantic analyzer to automatically select this sense of *of* over a dozen other productive senses of this preposition if this extra information were not provided. Typical N+PP collocations like these are often realized in text as NN compounds, as in the example *access to X* (access to the garage) vs. *X access* (garage access).

In our pre-evaluation corpus analysis, we found that leveraging lexically recorded PPs has quite high predictive power if the syn-struc contains only one PP, but that predictive power drops significantly if more than one PP is recorded, especially if the semantic constraints on the objects of the prepositions overlap. For example, a nominal sense of *training* includes the optional PPs *of X* and *by Z* to cover inputs like *the training of the athletes by the coaches*. If, however, we try to

leverage this PP-oriented information to automatically analyze nominal compounds, the question is, does the compound refer to the *of X* PP or the *by Z* PP? Is *citizen training* "training of citizens" or "training by citizens"? World and/or contextual knowledge is needed to make this distinction. However, on the positive side, the recording of multiple PPs can effectively narrow down the choice space of interpretations and, in the case of non-overlapping semantic constraints on prepositional objects (e.g., if an *of-PP* requires an abstract object whereas a *by-PP* requires a human object), the choice can be readily made by the analyzer.

Line 11. The next step in analysis represents a big leap in removing constraints on the interpretation process: the OntoSearch engine (Onyshkevych 1997) computes the ontological distance between pairwise combinations of all senses of N1 and all senses of N2 and posits the highest scoring result as a candidate analysis. For example, if two senses have nothing in common – e.g., PIG TAILPIPE – the path between the concepts will be long, indirect, and incur a high cost. By contrast, if the concepts can be linked by a single relation, then the nature and specificity of that relation is important. For example, say the analyzer encounters the NN *hospital physician*. OntoSearch will detect the following ontologically recorded statement:

PHYSICIAN
                    LOCATION   default   DOCTOR-OFFICE, HOSPITAL

Since HOSPITAL fills the default facet (i.e., it is a narrow constraint), *hospital physician* can be confidently analyzed as PHYSICIAN LOCATION HOSPITAL.

Of course, there are actually a lot of ways in which HOSTPIAL and PHYSICIAN could be linked using ontological search: since a hospital is a PLACE and a physician is a HUMAN, and HUMANs go to PLACEs, then the physician could be the AGENT of a MOTION-EVENT whose DESTINATION is HOSPITAL. Similarly, since a hospital is a PHYSICAL-OBJECT and since people can DRAW practically any PHYSICAL-OBJECT, then the PHYSICIAN could be the AGENT of a DRAW event whose THEME is HOSPITAL. The list of such analyses could go on and on. But the point is this: use of an essentially elliptical structure like a NN compound requires that the speaker give the listener a fighting chance of figuring out what he is talking about. Using the compound *hospital physician* to mean a hospital that a physician is drawing is simply not plausible. That lack of plausibility is nicely captured by ontological distance metrics: the closer the ontological distance, the more cognitively-motivated

the semantic correlation.

Now, one could argue that PHYSICIAN LOCATION HOSPITAL is not the most semantically precise analysis possible, which is true! If we wanted a better analysis, we could create a pattern that expected LOCATION followed by SOCIAL-ROLE, which would output the following types of meaning representations:

HUMAN		
	HAS-SOCIAL-ROLE	N2
	AGENT-OF	WORK-ACTIVITY-1
WORK-ACTIVITY-1		
	LOCATION	N1

Using this pattern, *hospital physician* would be superbly analyzed as a physician who works in a hospital; *bakery chef* would be equally superbly analyzed as a chef who works in a bakery, and so on. The point is that the agent will only attempt the unconstrained ontology-based reasoning of Line 11 if it failed to reach an analysis using the more narrowly guided approaches attempted in previous steps.

Line 12. If ontological distance measures do not result in a high-confidence contextual interpretation of the compound, then the system concentrates on just the meaning of the head, N2. The head, as the main contributor to the overall meaning of the clause, is objectively more important than the modifier represented by N1.

Line 13. If the head can be disambiguated, and if the meaning of N1 can be determined with relatively high confidence using the previously launched reference resolution procedures, then the agent analyzes NN as a generic RELATION between the referentially-disambiguated meaning of N1 and the contextually-disambiguated meaning of N2. This would happen in our *dog bed* example.

Line 14. If, by contrast, N2 can be contextually disambiguated but N1 cannot be referentially disambiguated, then machine learning can optionally be launched to try to establish an interpretation for N1, using the meaning of N2 as an input parameter. For example, if the agent was trying to select a meaning for *dog* in *dog bed*, and if it found corpus evidence of *cat bed* and *hamster bed* but no *human/person bed* (which reflects the result of a quick Google search), then it could guess that the ANIMAL meaning – i.e., DOG – was a better fit than either of the other two meanings – HUMAN (derog.) or FOLLOW. This is an interesting example because even though beds for humans are the default kinds of beds in the world, they are practically never referred to as *human beds* or *people beds*.

Line 15. If machine learning either isn't launched or does not result in a strong vote for the meaning of N1, then the agent can either (a) accept residual ambiguity, outputting the set of all possible analyses of N1 linked by the generic RELATION to the selected sense of N2 or (b) use an application-specific recovery strategy, such as asking a human collaborator what he means.

This ends the treatment of compounds in which both nominal elements are known words (i.e., some sense of them is present in the lexicon). The rest of the algorithm treats cases in which one or both nominal elements is not known. Processing can optionally involve machine learning using bootstrapping techniques with which we have experimented (e.g., Nirenburg et al. 2007). Although our experimentation is in the early stages, we believe machine learning methods could have high payoff potential, both for semi-automatic knowledge acquisition prior to system runs and for just-in-time support for processing unrecorded lexical stock.

Line 16. If the lexicon contains sense(s) of N2 but not N1, and if the agent can confidently disambiguate N2 (**Line 17**), and if machine learning for N1 is not available (**Line 18**), then the agent resolves the meaning as uninterpreted-N1 RELATION N2-INTERP – a result that will be sufficient to support overall clause-level semantic analysis despite the lack of understanding of the N1 modifier.

Line 19. If, by contrast, machine learning (ML) is available, then the agent can attempt to learn the meaning of N1, using the meaning of N2 as an input parameter. Imagine that the compound *bocci season* was encountered in the sentence *Bocci season begins in January and ends in March*, and that the OntoAgent lexicon lacked an entry for *bocci*.[6] Clause-level disambiguation of *season* would prefer the meaning TEMPORAL-SEASON over the event interpretation ADD-FOOD-SEASONING since the duration of the latter is measured in seconds, not months. OntoAgent could then search for corpus examples of NNs whose N2 was *season* in the meaning TEMPORAL-SEASON.

Let's start with the eventuality that the only examples it found included N1s indicating sports (this is not realistic in the general case but could occur in a domain-specific application). If the examples found were, exhaustively, *baseball season, hockey season, basketball season* and *golf season*, then the agent would recognize that, in one interpretation of each of these N1s, they are all children of SPORTING-EVENT; the agent could then posit, with high confidence, that *bocci* referred to

---

[6]Although we don't know of any universally agreed upon bocci season, a particular bocci club could presumably have a prime season.

some SPORTING-EVENT. Furthermore, it could posit a high-confidence relation linking TEMPORAL-SEASON and SPORTING-EVENT – namely, TIME-OF: TEMPORAL-SEASON TIME-OF SPORTING-EVENT. If the agent were tasked with broad-based knowledge acquisition, it could additionally try to learn ontological properties of bocci by reading and semantically analyzing texts that include the word, as reported in Nirenburg et al. 2007.

Another ML outcome is that the system finds examples whose N1 do not share an ontological parent: e.g., it might find *hunting season, fishing season, baseball season, concert season, opera season, duck season* and *deer season.* It could cluster these compounds based on their nearest ontological ancestor, generating the clusters: [DUCK-ANIMAL DEER] (descendants of ANIMAL), [CONCERT OPERA-EVENT] (descendants of ENTERTAIN-EVENT), and [HUNTING-EVENT FISHING-EVENT BASEBALL-GAME] (descendants of SPORTING-EVENT). The agent would have no reason to prefer any of these analyses over the others, and therefore the best it could do would be to analyze N1 as the set [or ANIMAL, ENTERTAIN-EVENT, SPORTING-EVENT] and put this set in a generic RELATION with the meaning of the head.

Line 20. If ML is attempted but no high-confidence results are achieved, then the agent must back off to either outputting [uninterpreted-N1] RELATION [interpreted-N2] or using an application-specific recovery strategy, as in Line 15.

Line 21. The agent reaches this step if it has determined that it cannot confidently disambiguate N2, and N1 is an unknown word. It considers the NN as a whole and attempts machine learning using analogical reasoning based on remembered instances of previously analyzed compounds.[7] (This type of reasoning could optionally be incorporated as heuristic evidence into earlier stages of processing as well.) For example, if the input is *tarsier bed* and the agent has a stored memory of *cat bed, dog bed* and *mouse bed* being confidently disambiguated as BED-FURNITURE LOCATION-OF SLEEP (EXPERIENCER *the animal in question*), then it can hypothesize that *tarsier* is an ANIMAL and create a corresponding semantically underspecified meaning representation. However, if it also had memories of *rose bed* and *flower bed*, then it would have to create a competing hypothesis that *tarsier* could be a kind of FLOWER.

Line 22. If ML does not offer a confident analysis, then the agent can back off to either outputting [uninterpreted-N1] RELATION [set of candidate interpretations of N2] or using an application-specific recov-

---

[7]Cf. the psycholinguistically-oriented work of Tagalakis and Keane 2005.

ery strategy, as earlier. This ends processing of NNs in which N2 is a known word.

Line 23. The agent reaches this point only if N2 is an unknown word – a more difficult situation than N1 being unknown since N2 is the head of the compound, whose meaning has to be integrated into the larger context.

Line 24. If N1 is known, and if reference-oriented disambiguation can be carried out with high confidence, then ML can optionally be launched, using the meaning of N1 as a input parameter. (Cf. Line 19.) If this leads to a high-confidence analysis, that analysis is used.

Line 25. Otherwise (if N1 is not known, or if it does not help with ML), the agent can attempt ML of the compound on the whole, as it did in Line 21.

Line 26. If analogical ML for the whole NN does not offer a confident analysis, then the agent can either employ whatever analysis has the highest score, or resort to an application-specific recovery strategy.

This ends the description of what we believe is the optimal approach to the analysis of nominal compounds by intelligent agents.

5   Evaluation

The algorithm described above is devoted to a full-scale solution for the problem at hand, not being constrained to one or a few of the large inventory of specific problems encountered when processing nominal compounds. The theoretical and descriptive work required to develop such a comprehensive approach must precede implementation work. There are many reasons for this conclusion, including the realization that without such prior work it would be problematic to integrate the treatments of individual subproblems while retaining good coverage and attaining high efficiency.

From the standpoint of evaluation, however, it is impossible to implement and evaluate such a comprehensive program of work all at one go, or to report all aspects of it in one paper. Therefore, our evaluation will cover a subset of the algorithm. While it is a subset, it is a non-toy, non-trivial subset. The main purpose of this evaluation is to determine if our approach is on the right track and to suggest possible correction courses and areas to focus our future development efforts.

The evaluation focuses on lexical and ontological patterns which we hypothesized would have high predictive power. This means that if two nouns can be interpreted using the expectations encoded in a listed pattern, then it is likely that they should be interpreted using that pattern. For example, the nouns in the NN *bass fishing* are ambiguous,

a subset of their meanings being presented below:

bass	*fishing*
BASS-FISH	FISHING-EVENT
STRING-BASS-INSTRUMENT	SEEK

Combining these meanings leads to 4 interpretations, paraphrased as:

1. carrying out the sport/job of fishing in an attempt to catch a type of fish called a *bass*
2. carrying out the sport/job of fishing in an attempt to catch a stringed musical instrument called a *bass*
3. seeking (looking for) a type of fish called a *bass*
4. seeking (looking for) a stringed musical instrument called a *bass*

However, only one of these interpretations, the first, matches a recorded NN pattern: FISH + fishing > FISHING-EVENT THEME FISH.

By analyzing *bass fishing* according to pattern 1, the system simultaneously selects a meaning of *bass*, a meaning of *fishing* and the relationship between them. The existence of this pattern asserts a preference for interpretation 1 as the default. We must emphasize, this is still only a default interpretation that needs to be incorporated into the clause-level semantic dependency structure during actual text processing. In this evaluation, we assessed how often this default interpretation was correct.

The evaluation corpus included texts from the 1987 Wall Street Journal.[8] We automatically extracted the evaluation subcorpus using a process that is not part of standard OntoAgent processing because it would have been unrealistic to run the entire corpus through the syntactic and semantic analysis engines to find the relatively infrequent examples that were within purview of this evaluation. So, we designed a method that would allow us to first string-search the corpus for candidate contexts that might be of interest, then prune those contexts to the actual evaluation corpus using a combination of automatic and manual processing.

Specifically, for each of the five classes of lexical and ontological patterns described above, we did the following, using the generate and test methodology (line numbers are from the algorithm):

1. (line 2) All N_N head entries in the lexicon were used, along with their plural forms, if applicable: e.g., phone_call, phone_calls.[9]

---

[8]The choice of corpus was irrelevant for our purposes.

[9]We generated plurals for this group but not for the others, due simply to experi-

2. (line 8, subclass 1) For all entries in which N2 was the head and fillers for N1 were listed in the syn-struc, all combinations of N1 N2 were generated: e.g., *I myself; degree Centigrade.*

3. (line 8, subclass 2) When N2 was listed as the headword and N1 was described using semantic constraints, we first generated a list of words in the lexicon that matched the given semantic constraints for N1. For example, returning to our fishing example, for the pattern FISH + fishing we generated a list of all words/phrases that mapped to the concept FISH or any of its ontological descendants; this yielded such strings as *bass, trout, tuna*, and so on.[10] Each of these was then combined with the headword to yield actual NN combinations, like *bass fishing* and *trout fishing*, which were searched for in the corpus.

4. (line 9) For all patterns comprised of two semantic constraints (i.e., a constraint on N1 and a constraint on N2), we generated a list of all words meeting N1's constraints and all words meeting N2's constraints. Then we combined all N1 strings with all N2 strings.

5. (line 10) When one N was listed as the headword and the other could be analyzed using a transformation from a PP construction, we generated a list of actual words that would match the semantic constraint on the object of the preposition. An example from earlier was "chain of X", which is analyzed as SET MEMBER-TYPE X, with the constraint on X that it refer to STORE or RESTAURANT. Accordingly, we generated a list of all words in our lexicon that refer to STOREs and RESTAURANTs. Then we combined these with the headword to create actual NNs to be sought in the corpus: *supermarket chain, restaurant chain, drugstore chain,* etc.

Using this methodology, we generated 53 million candidate search strings before halting the process. This did not exhaust the lexicon's potential for candidate generation but seemed sufficient for our goals of measuring the precision of our patterns. We used this candidate list to string-search the corpus. All hits were then parsed by the Stanford parser (de Marneffe et al. 2006).

---

mental oversight. The lack of plurals makes no difference to calculations of precision, which is what we are measuring here.

[10]We have not yet pursued automatic lexical expansion using, e.g., the hyponyms relation in WordNet. This would greatly expand our inventory of search strings and, presumably, our coverage. However, it is not without a downside: completely automatic expansion of this type would be error-prone due to homonymy, since the correct sense of a word must be used as a seed for expansion; semi-automatic expansion, for its part, is labor intensive.

In order to focus the evaluation on NN analysis separately from full system evaluation (which would involve glassbox analysis of all errors starting from preprocessing though syntactic analysis and including all aspects of semantic and pragmatic analysis), we included in the evaluation corpus only those candidate contexts that met all of the following criteria.

· They could be successfully parsed (i.e., throwing no errors) by the Stanford preprocessor and syntactic dependency parser.

· The Stanford parser recognized our target NN string as a compound.

· The NN was a 2-component compound, not part of a compound containing 3 or more nominals.

· The NN was not a proper noun or part of a proper noun.

· The OntoAgent semantic analyzer was able to process the Stanford output with no technical failures.

· The OntoAgent lexicon included at least one sense of each main word in the NN's clause (verb and arguments), since machine learning was not incorporated into this evaluation.

· The OntoAgent analysis of the verb and the NN was headed by an ontological concept rather than a modality frame, a call to a procedural semantic routine, or a pointer to a reified structure. This constraint was imposed simply to make the evaluation effort reasonably fast and straightforward. That is, we presented to the evaluator excerpts from text meaning representations rather than the full meaning representation for each sentence, which can run to several pages. We are confident that this pruning did not lead to skewing of the overall evaluation results, which focus on precision rather than recall.

· The NN served as an argument of the main verb of the clause, which permits clause-level disambiguation using bilateral selectional constraints (this is relevant for part 2 of the experiment). If the NN was, e.g., located in a parenthetical expression or used as an adjunct, then disambiguation would rely much more heavily on reference resolution and extra-sentential context.

The candidate pruning described above was carried out automatically, but manual inspection was used as a supplementary check. For example, quite often the syntactic parse did not recognize 3-noun compounds; less frequently it incorrectly labeled as a compound a structure actually composed of a noun followed by a verb. After this pruning, 72% of the examples initially extracted were deemed within purview of the evaluation, resulting in 935 examples.

The evaluation was carried out by a graduate student (who implemented an evaluation interface and various summary functions) with spot checking by a senior developer. Table 3 summarizes some salient aspects of the evaluation. Column 2 indicates the number of unique NNs (and total NNs, which reflects repetitions) detected for each analysis strategy. The evaluation orients around unique NNs because each NN ended up being either always analyzed correctly or always analyzed incorrectly. This reflects the fact that this evaluation is targeting only highly predictive patterns; more variability is expected of less constrained NN resolution strategies found later in the algorithm. In short, our approach does not deserve extra credit for getting *supermarket chain* correct 16 times, nor does it deserve to be unduly docked for getting *market share* wrong 125 times.[11]

TABLE 3 Evaluation Results

Strategy	Uniqe Exs. (of Total Exs.)	Correct, Incorrect, Residual Ambiguity
1	32 (of 223)	32 (100%), 0, 0
3	100 (of 455)	66 (66%), 10, 24
4	28 (of 204)	14 (50%), 9, 5
5	15 (of 53)	10 (67%), 3, 1

A small sampling of examples that were always correct, divided by strategy is shown in Table 4.

The "incorrect" analysis statistics show the number of NNs for which the only or multiple candidate analyses were all incorrect. Below are some examples of incorrect analyses, which are prefixed by the strategy that treated them and supplied with explanatory comments, as applicable (due to the length of the latter, these are not presented in tabular format). Multiple candidate analyses are indicated by a vertical bar. In the vast majority of cases, the NNs in question should be recorded as non-compositional phrasals.

> 3: *talk program* was analyzed as SOCIAL-EVENT PURPOSE (CONVERSATION | LECTURE), meaning a social event whose purpose is either conversation or lecturing (as might be plausible, e.g., as an activity for nursing home residents to keep them socially active). The intended meaning was a radio or TV program that involves talking rather than, say, music or drama.
>
> 3: *college education* was analyzed as TEACH THEME COLLEGE, meaning teaching about college.

---

[11]No corpus examples were found for NNs covered by analysis strategy 2.

TABLE 4 Sample Correct Analyses

Strategy	String	Analysis
1	telephone call	PHONE-CONVERSATION
	aircraft carrier	AIRCRAFT-CARRIER
	bulletin board	BULLETIN-BOARD
3	restructuring program	SOCIAL-EVENT PURPOSE REORGANIZATION
	exploration program	SOCIAL-EVENT PURPOSE INVESTIGATE
	guest program	SOCIAL-SERVE BENEFICIARY GUEST
	lung cancer	CANCER LOCATION LUNG
	neck injury	INJURY LOCATION NECK
4	supermarket chain	SET MEMBER-TYPE SUPERMARKET
	limousine fleet	SET MEMBER-TYPE LIMOUSINE
	customer satisfaction	SATISFY THEME CUSTOMER
	union member	SOCIAL-OBJECT MEMBER-OF LABOR-UNION
5	avocado tree	TREE PRODUCER-OF AVOCADO

3: *public education* was analyzed as TEACH THEME SOCIETY, meaning teaching about society.

3: *pilot program* was analyzed as SOCIAL-SERVE BENEFICIARY PILOT, meaning a social event that benefits airplane pilots, which is actually plausible but not intended in the corpus examples.

4: *home life* was analyzed as USABLE-LIFE DOMAIN DWELLING, meaning the length that a dwelling would be usable, e.g., what a contractor might think about, though one wouldn't normally say it like this in English.

4: *intelligence source* was analyzed as PLACE SOURCE-OF SPY-ON | INTELLIGENCE | INTELLIGENCE-INFO, meaning a physical location that is the source of a spying event, the abstract concept 'intelligence', or information that is gathered by means of spying.

5: *face amount* was analyzed as AMOUNT DOMAIN FACE, meaning the amount of the animal body part 'face'. Of course, this is actually a technical term in the domain of finance.

The "residual ambiguity" statistics indicate cases in which the analyzer posited more than one candidate analysis during the initial "surfacy" stage of NN analysis, before all sentence-level semantic analysis

capabilities were leveraged. It is the results of this stage of analysis that were used for our manual evaluation: i.e., we asked the question, "Did the patterns supply the analyzer with an option that, when considered in the larger dependency structure, would be selected as correct?" (Whether or not the analyzer actually did select the correct one when analyzing the sentence involves a large number of factors that would take far more space to explain sufficiently.)

Below is a sampling of examples belonging to the "residual ambiguity" category for the various strategies, supplied with comments as applicable. Each example is prefixed by the strategy that treated it.

3: *basketball program* was analyzed as SOCIAL-EVENT PURPOSE BASKETBALL | BASKETBALL-BALL. BASKETBALL refers to the game, whereas BASKETBALL-BALL refers to the object one bounces and throws. So the program is either devoted to games (e.g., by providing youth the opportunity to play the game) or to balls (e.g., by providing balls free of charge to schools) or to a vague combination of the two (if you're providing balls to schools, it's presumably to permit the students to play the game).

3: *sex education* was analyzed as TEACH THEME SEX-EVENT | GENDER. The options refer to teaching about the act of sexual intercourse or the notion of gender, e.g., gender differences. The first reading is intended by the corpus examples, and this meaning is stable enough to warrant recording as a lexicalized phrasal.

3: *colon cancer* was analyzed as CANCER LOCATION COLON-PUNCTUATION | COLON | MONEY. The options are that the disease CANCER is located either in a punctuation mark, in the BODY-PART of an animal called the COLON, or in the type of money called 'colon'. Obviously, the BODY-PART reading is expected and the analyzer can easily arrive at this interpretation automatically since the recorded pattern specifies that N1 must be a BODY-PART. To put a fine point on it, the notion 'residual ambiguity' is only residual up to the point when the analyzer semantically checks the correlation of all components.

3: *oil spill* was analyzed as SPILL THEME OIL | COOKING-OIL. The corpus examples refer to a fuel oil spill in an ocean; however, in a different context, this NN could also refer to the spilling of cooking oil, as in a kitchen. It would be risky to lexicalize this as 'fuel oil spill' because, in our approach, lexicalized-phrasal interpretations get a very large scoring bonus over compositional interpretations, effectively excluding the latter.

4: *ship fleet* was analyzed as SET MEMBER-TYPE SPACE-VEHICLE

| SHIP. Outside of context, the ships in question could either be spaceships or ocean ships. In our corpus examples, ocean ships were intended, but other contexts could as easily refer to spaceships.

4: *team member* was analyzed as SOCIAL-OBJECT MEMBER-OF SPORTS-TEAM | SET (MEMBER-TYPE ANIMAL). A team can either refer to a sports team or to a team more generally understood, as a team of workers collaborating on a project or a team of oxen. Context-level disambiguation and reference resolution must support this decision.

4: *fire source* was analyzed as PLACE SOURCE-OF DISCHARGE-WEAPON | FIRE. This could be a place where a fire started or a place where the firing of weapons is occurring.

5: *cancer patient* was analyzed as PATIENT EXPERIENCER-OF CANCER | CANCER-ZODIAC. This is an easy case for the analyzer: the event CANCER is ontologically described as having the case-role EXPERIENCER-OF whose default constraint is MEDICAL-PATIENT. By contrast, the ontological object referring to the zodiac sign 'cancer' does not have any case-roles.

5: *strawberry crop* was analyzed as CROP | CROP-PLANT PRODUCER-OF STRAWBERRY This is another easy case for the analyzer: only CROP is ontologically defined using the property PRODUCER-OF; CROP-PLANT (which refers to the food itself) is not.

5: *mint julep* was analyzed as MIXED-DRINK HAS-OBJECT-AS-PART CANDY | MINT. This ambiguity is tricky to resolve without knowing the recipe for mint julep: either it contains the herb MINT or it contains mint-flavored candies.

Let us summarize some relevant findings of this evaluation as well as our corpus work leading up to it. In our pre-evaluation analysis, certain classes of compounds proved to be notoriously difficult to analyze even wearing our finest linguistic hats. A star example is compounds with the headword *scene*, such as *labor scene, drug scene, jazz scene*. Of course, one could posit an underspecified ontological concept to account for this meaning of *scene* with an underspecified RELATION that links that concept to the kind of scene being described, but that would just be passing the buck. In reality, these concepts are adequately described only by scripts of the Shankian type, a different script for each type of scene.

Even if NNs are not as semantically loaded as *scene* ones, many more than one might imagine should be lexicalized as fixed expressions. All of the NNs that we recorded as headwords were analyzed correctly, which suggests that our lexicalization criteria, which are rather strict, are appropriate.

Metaphorical usages of NNs are not uncommon, just like metaphorical usages of other types of multi-word expressions. E.g., in the following corpus example, both *rabbit holes* and *storm clouds* are used metaphorically: "He also alerts investors to key financial rabbit holes such as accounts receivable and inventories, noting that sharply rising amounts here could signal storm clouds ahead." In some cases, automatically detecting metaphorical usage is straightforward, as when the NN is preceded by the modifier *proverbial*: "'They have taken the proverbial atom bomb to swat the fly,' says Vivian Eveloff, a government issues manager at Monsanto Co. ..." In other cases, our expectations were countered by an unforeseen narrow-domain usage of a term: e.g., it turns out that *body part* can be used to refer to part of a vehicle.

Overall, our recorded patterns worked as expected (factoring out entities that are actually non-compositional) and we are encouraged that the approach of seeking high-confidence pattern-based analyses before resorting to less high-confidence analysis strategies is a good one.

6   Final Thoughts

This paper has concentrated on the analysis of two-noun compounds. However, we took pains to ascertain that the approach can be extended to treating larger compounds. Indeed, in this latter case the agent would first seek out islands of highest confidence (i.e., high-scoring 2-noun interpretations), then combine those partial analyses. Consider, for example, the 3-noun compound *ceiling height estimation*. The candidate interpretations are [[ceiling height] estimation] and [ceiling [height estimation]]. The first candidate analysis will receive a very high score using rule m of Section 3 for "ceiling height" (If N2 is PROPERTY and N1 is a legal filler of the DOMAIN of N2 then N2 DOMAIN N1) and rule j of Section 3 for "height estimation" (If N2 is EVENT and N1 is a "default" or "sem" THEME of N2 then N2 THEME N1). By contrast, the second candidate analysis will receive a much lower score because there is no high-confidence rule to combine CEILING with ESTIMATE (CEILING is not a sem or default filler of the THEME case-role of the event ESTIMATE). Although it would be unwise to underestimate the devil lurking in the details of processing large compounds, it is reasonable to assume that multi-noun compounds can be treated by an extension to the algorithm presented here rather than requiring a fundamentally different approach.

This algorithm for NN analysis is, like most algorithms targeting linguistic subphenomena in OntoAgent, being implemented in stepwise

fashion. The resources, processors and methodologies brought to bear are not unique to NN analysis, making this corner of work a variation on the theme of basic semantic analysis.

Although approaches to NLP that incorporate even a modicum of manual resource acquisition have been widely discounted over the past two decades due to concerns about the so-called "knowledge bottleneck", we consider this pessimistic assessment misplaced for several reasons. First, when NLP is approached as a singleton functionality, then the time and effort needed to build knowledge resources might seem excessive; however, when NLP is treated as one of many functionalities of an intelligent agent – and when the agent can reuse the same knowledge resources for all its functionalities – then the cost effectiveness increases dramatically. Second, although practitioners of knowledge-lean methods claim to be avoiding the knowledge bottleneck, there are numerous caveats: a) many of the more successful approaches require manually annotated (i.e., expensive) corpora for training; b) many of the single subproblem-oriented engines thus trained cannot perform well – or at all – in the absence of manually preprocessed input; and c) all past corpus annotation efforts cover only the simpler cases of any given phenomenon, thus leaving large expanses of language use outside of purview (cf. McShane and Nirenburg 2013). Third, knowledge-based approaches like those taken in OntoAgent are readily extensible to other languages, with significant potential for reuse of resources. Finally, since there is much overlap in the processing of all input – NN compounds, so-called "multi-word expressions", compositional clause-level dependencies, etc. – progress on any of these contributes to progress on all of them.

The purpose of this paper is four-fold: (1) to juxtapose a cognitively-oriented approach to nominal compound analysis with the more widespread mainstream NLP approaches; (2) to elucidate the actual scope of eventualities that an agent can encounter when processing NN compounds (in a similar spirit as is done for reference resolution in McShane 2009); (3) to share a concrete, top-down, implementable (at present still only partially implemented) approach to analyzing NN compounds that should be applicable to cognitive architectures outside of OntoAgent; and (4) to present our initial evaluation of the OntoAgent NN processing capabilities, which involves an attempt to answer the difficult question, *How can one fairly evaluate knowledge-based systems whose goals involve advancing our understanding of scientific issues over the long term, considering that the evaluation will not be equivalent to those used for near-term applications that exclude many eventualities?* This question, much discussed in conference corridors but resistant to a neat

or universally satisfying answer, deserves careful consideration if long-term research endeavors are to hold a place in the landscape of scientific investigation.

Acknowledgments

References

Beale, Stephen. 1997. *Hunter-Gatherer: Applying Constraint Satisfaction, Branch-and-Bound and Solution Synthesis to Computational Semantics*. Ph.D. thesis, Language and Information Technologies Program, School of Computer Science, Carnegie Mellon University.

de Marneffe, Marie-Catherine, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. LREC.

Downing, Pamela. 1977. On the creation and use of English compound nouns. *Language* 53(4):810–842.

Finin, Timothy. 1980. *The Semantic Interpretation of Compound Nominals*. Ph.D. thesis, University of Illinois.

Gagné, Christina L. and Thomas L. Spalding. 2006. Using conceptual combination research to better understand novel compound words. *SKASE Journal of Theoretical Linguistics* 3:9–16.

Hendrickx, Iris, Zornitsa Kozareva, et al. 2013. Semeval-2013 task 4: Free paraphrases of noun compounds. Seventh International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics.

Isabelle, Pierre. 1984. Another look at nominal compounds. Annual Meeting of the Association of Computational Linguistics.

Kim, Su Nam and Preslav Nakov. 2011. Large-scale noun compound interpretation using bootstrapping and the web as a corpus. pages 648–658. 2011 Conference on Empirical Methods in Natural Language Processing.

Lapata, Maria. 2002. The disambiguation of nominalizations. *Computational Linguistics* 28(3):357–388.

Levi, Judith N. 1979. The syntax and semantics of complex nominals. *Language* 55(2):396–407.

Lieber, Rochelle and Pavol Štekauer. 2009. *The Oxford Handbook of Compounding*. Oxford University Press.

McShane, Marjorie. 2009. Reference resolution challenges for an intelligent agent: The need for knowledge. *IEEE Intelligent Systems* 24(4):47–58.

McShane, Marjorie and Sergei Nirenburg. 2012. A knowledge representation language for natural language processing, simulation and reasoning. *International Journal of Semantic Computing* 6(1):3–23.

McShane, Marjorie and Sergei Nirenburg. 2013. Use of ontology, lexicon and fact repository for reference resolution in Ontological Semantics. In A. Oltramari, P. Vossen, L. Qin, and E. Hovy, eds., *New Trends of Research in Ontologies and Lexical Resources*, pages 157–185. Springer.

McShane, Marjorie, Sergei Nirenburg, and Stephen Beale. 2005. An NLP lexicon as a largely language independent resource. *Machine Translation* 19(2):139–173.

Moldovan, Dan, Adriana Badulescu, Marta Tatu, Daniel Antohe, and Roxana Girju. 2004. Models for the semantic classification of noun phrases. pages 60–67. HLT-NAACL 2004 Workshop on Computational Lexical Semantics.

Nirenburg, Sergei, Tim Oates, and Jesse English. 2007. Learning by reading by learning to read. International Conference on Semantic Computing.

Onyshkevych, Boyan. 1997. *An Ontological-Semantic Framework for Text Analysis*. Ph.D. thesis, Center for Machine Translation, Carnegie Mellon University.

Rosario, Barbara and Marti Hearst. 2001. Classifying the semantic relations in noun compounds via a domain-specific lexical hierarchy. pages 82–90. Empirical Methods in Natural Language Processing.

Tagalakis, Georgios and Mark T. Keane. 2005. How understanding novel compounds is facilitated by priming from similar, known compounds. 27th Annual Conference of the Cognitive Science Society.

ter Stal, Wilco G. and Paul E. van der Vet. 1994. Two-level semantic analysis of compounds: A case study in linguistic engineering. CLIN.

Tratz, Stephen and Eduard Hovy. 2010. A taxonomy, dataset, and classifier for automatic noun compound interpretation. Association for Computational Linguistics.

Vanderwende, Lucy. 1994. Algorithm for automatic interpretation of noun sequences. pages 782–788. COLING.