Proceedings of the International Multiconference on
Computer Science and Information Technology pp. 517–524
ISBN 978-83-60810-22-4
ISSN 1896-7094

Multiple Noun Expression Analysis:

An Implementation of Ontological Semantic Technology

Julia M. Taylor
RiverGlass, Inc &
Purdue University
USA
jtaylor1@purdue.edu

Victor Raskin
Purdue University &
RiverGlass, Inc
USA
vraskin@purdue.edu

Maxim S. Petrenko
Dashkova University &
RiverGlass, Inc
Russia & USA
mpetrenk@gmail.com

Christian F. Hempelmann
RiverGlass, Inc &
Purdue University
USA
kiki@riverglassinc.com

***Abstract*—The paper analyzes multiple noun expressions, or compound nouns, as part of the implementation of Ontological Semantic Technology, which uses a lexicon, an ontology, and a semantic text analyzer to access and represent the meaning of text. Because the analysis and results depend on the lexical senses of words, general principles of lexical acquisition are discussed. The success in interpretation and classification of such expressions is demonstrated on 100 randomly selected sequences of noun compounds.**

***Keywords*—multiple noun expressions, meaning interpretation, ontological semantic technology**

I. Introduction

THIS paper describes the implementation of Ontological Semantic Technology (OST), an offshoot of Ontological Semantics [1], a theoretical and computational approach to meaning in natural language. OST is an advanced, improved, and revised application of Ontological Semantics for real-life commercial systems with its latest implementation currently being developed for RiverGlass, Inc. We will illustrate the process of OST implementation by presenting the elements of its ontology and lexicon, which are activated to analyze the difficult semantics of noun + noun expressions, a specific case that the OST multiple word expression (MWE) module is responsible for handling, and comparing the output of the OST semantic analyzer with an informal taxonomy of the meaning relations between the constituent nouns in these constructions. This approach [2] is not dependent on any training corpus but rather on the semi-automated acquisition of the OST static resources (the lexicons and ontology) which can be used for any text. Based on these resources, all OST processing is fully automated. In order to evaluate whether sentences or expressions are interpreted correctly by the OST software, familiarity with OST ontology is required. While the final evaluation of the approach can only occur on the full implementation of an application, we describe partial metrics both to assess progress and to improve the resources and the software, as necessary.

MWEs are well-known as a notorious problem in NLP [3] and workshops have been dedicated to their analysis since 2003 (http://multiword.sourceforge.net/). The central tasks that MWEs pose are their identification (extraction) as being multiword and not co-occurrences of several single words, and the interpretation and representation of their meaning. Our meaning-based linguistic method focuses on the latter task, which subsumes the former by necessity. Common approaches to the MWE subclass of noun compounds proceed without wanting to use costly-to-acquire world knowledge that would distinguish *fish knife* from *steel knife* with the help of knowing that *fish* are edible and not used as the material of artifacts while *steel* is not edible, but used as material in artifacts. The only knowledge used in such approaches is contextual clues [4], ideally specific paraphrases of the compounds, typically used in supervised learning approaches, e.g., [5]. While the relation between the concepts represented by the nouns in compounds is infinite, these approaches postulate subsets of relations, e.g., [6], [7], into which they aim to classify the compounds, with smaller sets naturally resulting in better performance numbers and the items in the sets almost never being motivated by an application. These consequently very coarsely grained subsets (commonly four to twenty items, cf. [8]) are often mapped onto prepositional paraphrases, in which *fish knife* would become *knife for fish* and *steel knife* would become *knife of steel.* Our approach, on the other hand, allows any property or relation available in our ontology to hold between the concepts and aims at the correct identification of each property with its functors.

II. OST Resources

A. OST Ontology

The OST ontology attempts to capture the users' knowledge of the world in a language-independent way. Text is interpreted in terms of the knowledge in the ontology; thus, it is important that as many relationships among concepts as necessary are accurately captured.

Formally, the ontology is a lattice of logically structured concepts (for a formal representation see [9]). It is divided into EVENTS, OBJECTS, and PROPERTYS, with the first two further divided into PHYSICAL, MENTAL, and SOCIAL subcategories. EVENTS and OBJECTS are connected through PROPERTYS, while strictly adhering to inheritance rules, creating the required richness of interrelationship among concepts. Figure 1 shows the top levels of the OST ontology; Figure 2 shows a sample concept with its (non-inherited) properties.

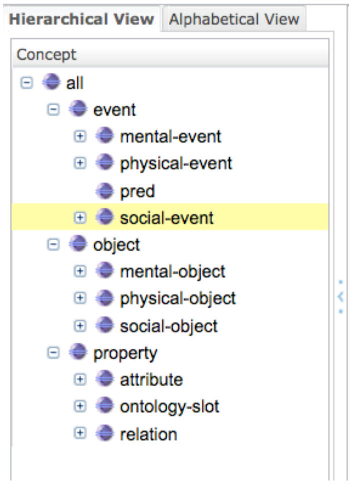

Figure 1. Top levels of OST ontology

B. *OST Lexicon*

The structure and format of lexical entries as well as lexicon acquisition strategies within ontological semantics have been discussed at length in [1: 230-245, 322-350]; acquisition techniques have been outlined in [10]; and the development of a toolkit for automated acquisition within OST has been proposed in [11]. This section will offer a concise outline of the structure and format and briefly discuss procedures of lexicon acquisition, only to the extent that they pertain to the paper.

1) Definition, functions, and format

In OST, all language-specific information is encoded in the lexicon. The lexicon is a machine-readable repository of words, non-compositional units (e.g., "hot dog", "gut feeling") and idiomatic expressions (e.g., "buy the farm"), whose meaning is defined through ontological concepts and whose morphology and syntax is defined through a set of part-of-speech tags, syntactic role tags and syntactic variables, properly ordered and co-indexed with their semantic counterparts.

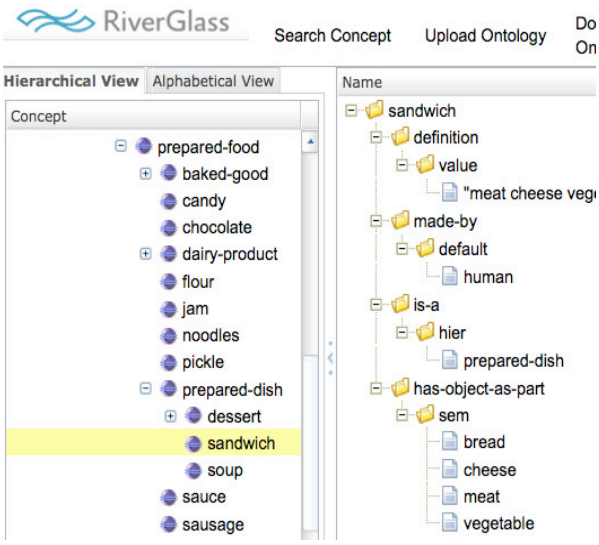

Figure 2. The concept SANDWICH

In relation to the ontology, the content and structure of which is language-independent and parsimonious, the lexicon captures all linguistic idiosyncrasies including language-specific senses, non-literal extensions, synonymy, homonymy, grammatical derivatives, optional phrasal particles, etc. As a resource immediately accessible by the Semantic Text Analyzer (STAn, see Section II C) during processing, the lexicon (1) directs the analyzer to the appropriate ontological concept, its properties, and their fillers, so that further computation can be performed based on that concept's definition, and (2) follows restrictions stipulated in the ontology so that property fillers in lexicon senses can only narrow down (i.e., be children of) those outlined for respective ontological concepts or introduce new properties or fillers outside the branch of the parent's filler's ancestor.

A lexical entry template is shown below:

```
(head-entry
    (sense-1, 2, 3…
        (cat(n/v/adj/pro/prep))
        (synonyms "")
        (anno
            (def "")
            (comments "Acquired by <acquirer name> on <date>
                at <time>. ")
            (ex "")
        )
        (syn-struc(
            (root($var0))(cat(n/v/adj/pro/prep))
            (subject/object((root($var#))(cat(np/vp/s))))
            )
        (sem-struc
            (root-concept
                (property(value(^$var#
                (should-be-a(default/sem(concept))))))
        )
    )
)
```

The fields "synonyms", "anno", "comments", and "ex" are of no value to the computer and serve the human acquirer. The fields "cat", "syn-struc", and "sem-struc", meaning "category", "syntactic structure", and "semantic structure", respectively, contain crucial information about the sense used by the analyzer in meaning computation.

The syntactic structure captures the position(s) a word can take in a clause: the root variable locates the word itself, and, for the case of events, syntactic roles of subject or object are indicated, along with their categorical features. For every syntactic role, a variable in the root stands for a specific word carrying this function. Every new syntactic role is assigned a new variable. If word order variation is possible, additional syntactic structures are listed.

The semantic structure anchors the meaning of the sense in ontological concepts, commonly a head concept most defining for the sense. In the case of EVENTs, every syntactic role indicated in the syntactic structure is given its case role interpretation drawn from the ontological definition of the respective EVENT. Variables for each case role are co-indexed with their counterparts in the syntactic structure. Fillers for case roles and other properties tighten more general ontological restrictions where necessary via the SHOULD-BE-A relation, the purpose of which is to further specify the ontological nature of the variable, and the DEFAULT or SEM facets, which prioritize and assign weights to the property values [9].

Figure 3 shows the entry "hot-dog-n2", meaning a sandwich with a sausage rather than the other common sense of "hot dog" which is just the sausage by itself:

```
(hot-dog-n2
   (cat(n))
   (synonyms "")
   (anno
      (def "a sandwich with a frankfurter in a split roll")
      (ex "he ordered a hot dog"))
   (syn-struc((root($var0))(cat(n))))
   (sem-struc(sandwich(contains(sem(sausage)))))
)
```

Figure 3. Lexicon sense of hot-dog

Since no direct concept for a hot dog is available in the ontology, the sense has been acquired by finding the nearest available concept SANDWICH and constraining it with the property CONTAINS filled with the concept SAUSAGE. All ontological restrictions have been observed and further specified: in the ontology, the domain and range of the property CONTAINS are filled with the concept PHYSICAL-OBJECT (of which SANDWICH and SAUSAGE are descendants).

2) *Lexicon acquisition*

A standard acquisition procedure implies creating machine-readable descriptions of lexical senses by listing appropriate concepts and their properties and restricting their fillers where needed. Given that defining the number of lexicon senses, choosing the appropriate head concept, and determining the "tightness" of restrictions for property fillers requires human competence, lexicon acquisition is largely human-driven. However, certain aspects of acquisition (entry format, inherited property fillers, domain/range compatibility check, template-based acquisition of specific word classes, etc.) lend themselves to automation, as demonstrated by various implementations of OST.

Depending on application objectives, the state of completion of the ontology, and the functionality of the Semantic Text Analyzer, three acquisition techniques have proven useful in various implementations [10]. Ontology-driven acquisition leads to a lexicon with the minimal lexicalization of concepts (i.e. one entry per concept), which is time-efficient but shallow and can only be used at early phases of ontology acquisition. Domain-driven acquisition focuses on topic-related vocabularies and onomasticons, repositories of named entities, as well as specific word classes, and is aided by the analysis of resulting text-meaning representations (TMRs) of domain-specific corpora and "gap detecting" toolkits identifying missing senses with further part-of-speech sorting. Analyzer-driven acquisition, the main method for quality-control and improvement, implies testing every newly acquired entry by the analyzer, followed by TMR analysis and subsequent modifications of the lexicon entry, ontological concepts, or functioning analyzer modules. The approaches are supplementary to each other, although, in practical applications, the domain-driven acquisition tends to dominate in order to ensure optimal lexical coverage and to minimize unattested input.

C. *Semantic Text Analyzer and InfoStore*

The Semantic Text Analyzer (STAn) is a software that interprets text according to knowledge from ontology and lexicon and assigns a TMR to each clause, with textually related clauses as linked TMRs. STAn does not have a graphical interface as its output is written directly into the InfoStore database, which collects all correctly processed TMRs. Thus, InfoStore is in unique possession of all machine-understandable interpretations of all texts processed, which it can then use to guide further interpretations by adding the counterpart of the linguistic and extralinguistic knowledge that humans use to disambiguate, complement meaning, and reasoning contextually.

For the work reported in this paper, we have not yet used InfoStore, but concentrated on STAn's interpretation results obtained only with the current limited ontology (about 2,000 concepts) and English lexicon (about 25,000 senses). The influence of the InfoStore on the overall interpretation process and the treatment of MWEs in particular as well as the resulting ontological adjustments deserve treatment in a separate paper.

Figure 4 shows STAn's interpretation of the sentence *management attempts to comply with local employment legislation* in debug mode. A more readable version of the final TMR is in Figure 5.

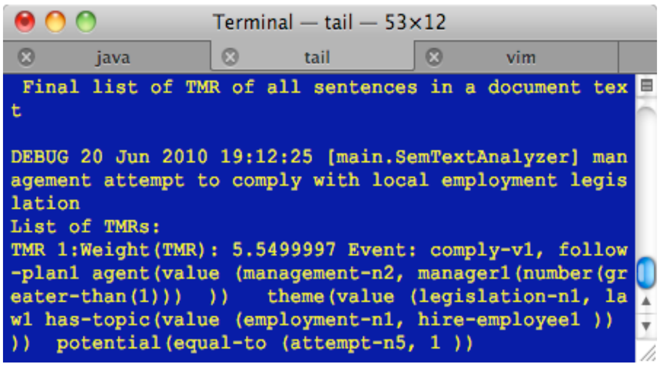

Figure 4. STAn's interpretation of *management attempts to comply with local employment legislation*

```
follow-plan1
   (agent(value(manager1
      (number(greater-than(1)))
   )))
   (theme(value (law1
      (has-topic(value (hire-employee1)))
   )))
   (potential(equal-to (1)))
```

Figure 5. A formatted TMR of *management attempts to comply with local employment legislation*

Figure 6 shows results that are derived from processing done with a simple InfoStore to illustrate the capabilities of an

OST system. It shows a set of documents retrieved from InfoStore that use the concept for firing of employees, listed separately as to which surface word triggered the concept, e.g., *fire*, as opposed to *terminate*, *discharge*, *dismiss*, etc.

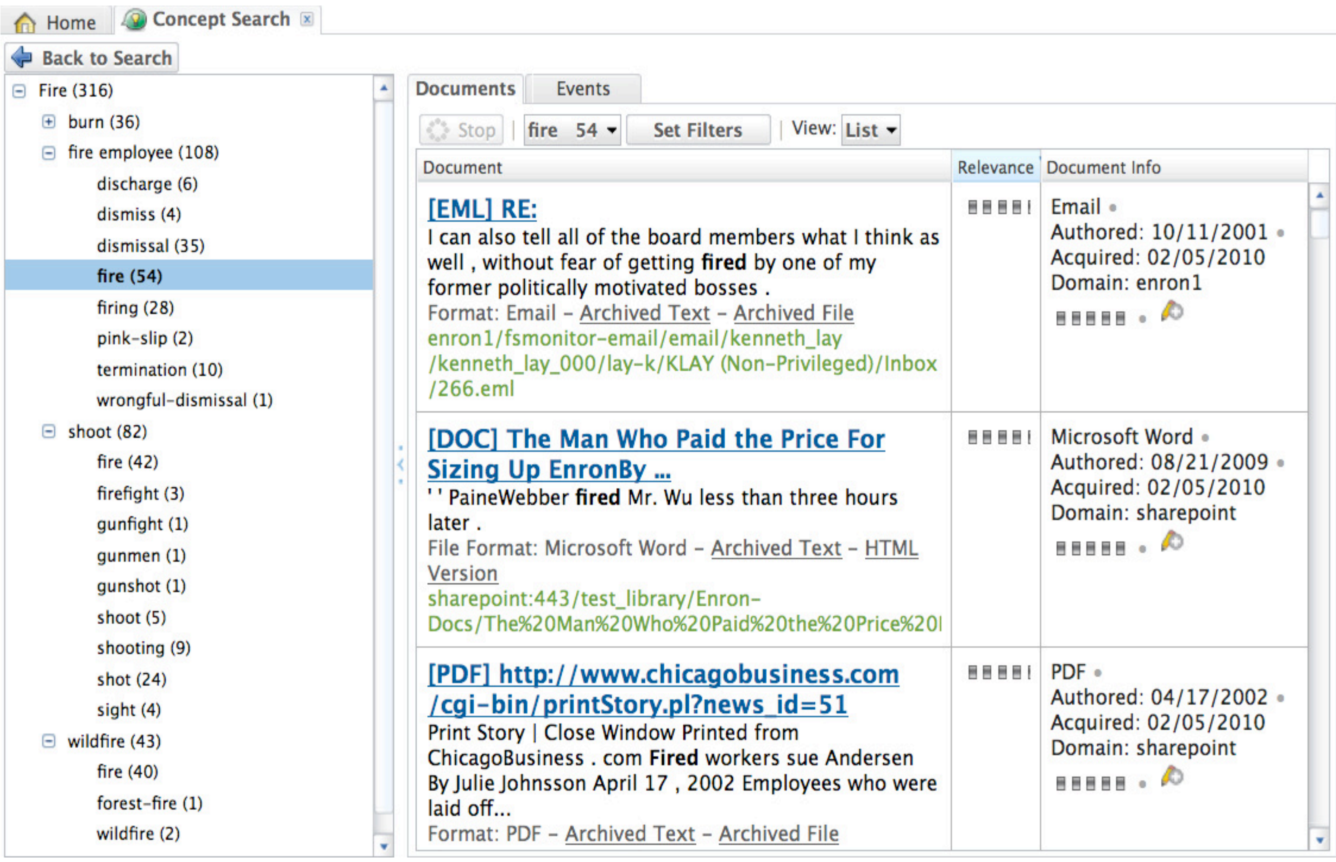

Figure 6. An example of processing with InfoStore

III. Multiple Word Expressions: N+N[+]

In this section, we apply OST to the semantic analysis of English compound nouns. Such constructions are notoriously difficult to interpret, not only because of the English morphology which allows the adjectival usage of the first noun but mostly because any such construction can be analyzed syntactically as the transform of a clause, where the two nouns can be in any number of meaning relations. In other words, syncretism, prevalent in natural language in general, manifests itself here in the worst possible ways.

For example, the construction *IBM lecture* can mean a lecture about IBM, by IBM, at IBM, or sponsored by IBM ([1] see also [4, 12]), and the human hearers/readers have to use their knowledge of the world to understand the construction correctly—or to ask a clarification question about the intended meaning. Since no syntactic analysis can differentiate among the various possible semantic relations between the two words of a noun + noun compound, it is a good test case for the OST semantic capabilities.

It appears that [13: 589-591] contains a reasonably grain-sized listing of the possible meaning relations, which is reproduced below in an abbreviated form:

- Composition
  - N2 is made from N1
  - N2 consists of N1
- Purpose
  - N2 is for the purpose of N1
  - N2 is used for N1
- Identity
  - N2 has the same referent as N1 but classifies it in terms of different attributes
- Content
  - N2 is about N1
  - N2 deals with N1
- Source
  - N2 is from N1
- Objective type 1
  - N1 is the object of the process described in N2 or of the action performed by the agent described in N2
- Objective type 2
  - N2 is the object of the process described in N1
- Subjective type 1
  - N1 is the subject of the process described in N2
  - N2 is nominalized from an intransitive verb

- Subjective type 2
  - N2 is the subject of the process described in N1
- Time
  - N2 is found at the time given by N1
- Location Type 1
  - N2 is found or takes place at the location given by N1
- Location Type 2
  - N1 is found or takes place at the location given by N2
- Institution
  - N2 identifies an institution for N1
- Partitive
  - N2 identifies parts of N1
- Specialization
  - N1 identifies an area of specialization for the person or occupation given in N2; N2 is animate

It is immediately disclaimed in [13] that some cases may belong to two or more classes and that some cases may belong to none. What is clear is the mixed syntactic/semantic nature of the taxonomy, with the objective- and subjective-types clearly standing out as shallow and, therefore, semantically vague, thus guaranteeing the pretty crippling heterogeneity of the members of these classes.

We have selected the first 100 noun + noun expressions from the Enron corpus (http://www.cs.cmu.edu/~enron/), starting from a randomly selected text, and manually assigned each of them to the most appropriate of the class(es) above. We excluded from the selection the multi-word phrasals that OST treats as single lexicon entries as well as proper nouns. Several of the expressions were of the noun + noun + noun type, and we considered them as a single expression rather than two: the selected 100 expressions contain 6 triple-noun ones. Next, we ran STAn on the actual sentences containing these constructions and compared the OST treatment of them with the above taxonomy.

It is important to understand that OST deals with deep semantics, and therefore recognizes the fact that a shallow semantic—or even such a morphosyntactic category as part of speech—does not determine the ontological basis of a word. Thus, nouns in OST can be anchored in EVENTs (a walk), PROPERTYs (color), or OBJECTs, even though the last type is prevalent by far. The part of speech is used as a clue, often an unimportant one, along with other grammatical (morphological and syntactic) information because the emphasis in OST analysis is always on matches of semantic properties. These clues are nevertheless collected and used as needed in the analysis. As mentioned, noun + noun expressions are particularly interesting and difficult to analyze precisely because the grammatical clues fail to help in differentiating among the numerous types of connection between N1 and N2.

For any N1 + N2 construction, therefore, the Semantic Text Analyzer looks for a property of, typically, N2 such that N1 falls in the range of this property's value. The predominantly prepositive nature of the English adjectival constructions makes the reverse situation—when N2 is the value of the property of N1—rarer. The property, or semantic relation, is (automatically) selected by STAn so that:

$$\max_{p\in P}\{\max_{fct\in Facet}\{I_C(p(fct(I_D)))*\sum_i const*inhd(i)\}\},$$

where P is the property (p) set, Facet is the facet (fct) set, and I_C and I_D are interpretations of the meaning of N_1 and N_2 with C and D being concepts in $\mathcal{D}$; where $\mathcal{D}$ is the disjoint union of $\mathcal{D}$c (concepts) and $\mathcal{D}$d (literals), and given its interpretation function $\mathcal{I}$, for every atomic concept B, $\mathcal{I}$[B] $\subseteq\mathcal{D}$c; $\mathcal{I}$ [C(Rel(Facet(D)))] $\subseteq\mathcal{D}$c, $\mathcal{I}$ [C(Rel(Facet(D)))] = $\mathcal{I}$ [C] $\cap\mathcal{I}$ [Rel(D)]; and where i is C, D or p and inhd(i) is the inheritance distance measure.

In other words, after STAn selects all the properties of the concept(s) underlying either noun, such that the meaning of the other noun fits the range of those properties or the properties themselves, the main task becomes the selection of the most appropriate connecting property, and this is where the weighting metric described above becomes crucial.

A. *Analysis of 100 N+N(+N) with Biber's taxonomy*

We agree with [13] that some of the N+N expressions fall under several categories: we found 17 such cases. Most of these due to somewhat vague definitions of such rubrics as Content and Identity, but others resulting from the mixed syntactic/semantic nature of the taxonomy, so that *business community planning* could be described equally well as Content or Objective type 1. Also, an expression could be ambiguous, and the alternative meanings could belong to different rubrics and the sentence containing such an expression was not sufficient for disambiguation in isolation (and, as we mentioned above, we are not using InfoStore for this paper).

Table I shows the taxonomic distribution of the 100 expressions; those that belonged to two or more rubrics were counted as many times, thus driving the total above 100. The rubrics not represented in our sample are omitted.

TABLE I. TAXONOMIC DISTRIBUTION OF THE 100 EXPRESSIONS

Taxonomic rubric	**Number of expressions**
Composition	1
Purpose	11
Identity	20
Content	22
Objective type 1	19
Objective type 2	3
Subjective type 1	15
Subjective type 2	3
Time	2
Location type 2	3
Institution	7
Partitive	4
Specialization	6
No appropriate rubric	2

This open-ended and somewhat vague classification is hardly suitable for computation. In the next section we will demonstrate the results of the OST analysis of the same expressions, and compare these results with these in the most populous rubrics above.

B. *Analysis of 100 N+N(+N) with OST*

As mentioned above, English nouns can be anchored in OBJECTs, EVENTs, or PROPERTYs from the OST ontology. Thus, a more meaningful, although still crude classification is between these types, shown in Table II.

In English adjectives predominantly precede the nouns that they modify. Accordingly, the adjectival use of nouns before other nouns, possible in English because of its impoverished morphology, makes N1s in N1 N2 sequence the modifier of N2 much more frequently than the other way around. This premodification prevails in our sample, with just a sample of postmodification expressions.

TABLE II. TOP LEVEL ONTOLOGICAL DISTRIBUTION OF THE 100 EXPRESSIONS

N1 N2 (N3)	**Number of expressions**
Event Event	11
Event Object	13
Event Property	4
Object Event	19
Object Object	33
Object Property	6
Property Object	5
Property Event	1
Object Event Object	3
Property Event Object	1
Property Object Event	1
Object Object Object	2
Object Event Event	1

STAn relies on premodification-oriented rules and goes into the postmodification regime only after failing to find any possible premodified interpretation. As an example, consider *monitor performance* (taken from our data): while it is likely that a human will interpret it as monitor for performance (postmodification), the premodified interpretation is performance of the monitor, and it is the latter that was easier for STAn to access according to the rules it uses.

Honoring, as it were, the predominance of premodification in the sample and in the language, we will refer to N1+N2 constructions as EVENT-, OBJECT-, or PROPERTY- driven when N2 is an EVENT, OBJECT, or PROPERTY, respectively. The interpretation of the EVENT-driven expressions is clearly delimited by the properties of the events, such as AGENT, INSTRUMENT, THEME, LOCATION, etc, and just as many objects in a regular sentence fit comfortably enough as the fillers of these properties, they do so in the N+N expressions. Unfortunately, this group is minority in the sample (33 instances). The EVENT + EVENT combinations present several additional alternatives for the analysis, and the interpretation is typically more difficult, often requiring InfoStore input.

The PROPERTY-driven compounds are easy to compute as the portion most difficult to find, the property, is explicitly stated in N2. The only work that is required is then to check whether N1 should be in the domain or range of the given property.

The OBJECT-driven compounds, which constituted the majority of the sample (57 instances), are more difficult to interpret correctly than the previous two groups. Problematic among them are OBJECT + OBJECT expressions that require some connection, where this connection might go far beyond a single property, but rather require a “story” involving the two objects. Ellipses (see below) are rampant in this category: thus, *group plan* is actually an insurance plan for the group.

Using the taxonomy in [13], Table III indicates significant OST subclasses in the most populous taxonomic rubrics where the nature and difficulty of computational treatment differ a greatly from one subclass to another.

TABLE III. TOP LEVEL ONTOLOGICAL DISTRIBUTION OF THE 100 EXPRESSIONS

Biber \ OST	Object=N2	Event=N2	Property=N2
Identity	15	3	2
Content	15	6	1
Objective 1	8	11	0
Subjective 1	1	12	2

It is interesting to see how many different connecting properties OST establishes on the OBJECT-driven subclasses, the two most populous ones being the Identity and Content rubrics. The properties listed here were found by STAn and recognized by human expert judges as correct:

- Identity/OBJECT-driven:
  - HAS-SOCIAL-ROLE
  - SALIENCY
  - RANK
  - EMPLOYED-BY
- Content/OBJECT-driven:
  - HAS-TOPIC
  - REPRESENTED-BY

Thus, the OST properties provide a finer grain size. They also can move an expression from one rubric to another, overriding the weak intuition of a human taxonomist or confirming one of the possible choices and excluding the other.

Let us repeat the procedure for the next two most populous EVENT-driven subclasses:

- Objective type 1/EVENT-driven:
  - INSTRUMENT

   - EXPERIENCER
   - THEME
   - THEME-OF
   - HAS-TOPIC
- Subjective type 1/EVENT-driven:
   - AGENT
   - LOCATION
   - THEME
   - EXPERIENCER
   - BENEFICIARY
   - PART-OF-EVENT

The 6 subtypes of the last type are particularly noteworthy because they include almost all of the common event properties in OST, INSTRUMENT notoriously but understandably missing.

C. Evaluation

1) Types of evaluation

There are multiple factors contributing to the process of correct interpretation in OST. As far as noun compounds are concerned, these factors include:

- First and foremost, the correct disambiguation of each noun's meaning.
- The identification of the possible connection(s) between the nouns in the expressions, according to the ontological world model.
- The discovery of the best interpretation of the expressions.
- The assignment of the highest rank (weight) to the TMR with the best interpretation.

2) The numbers

For the final version of the paper, we used STAn 4.0 for the evaluation. In the 100 expressions that we used, there were 205 words, most of them with several senses in the OST lexicon. 196 senses of them (or 95.6%) were selected correctly. In the few cases when the words had only one sense each in the lexicon, OST still checked (and the human expert verified) whether the sense fit. So, technically, in 95.6% of the nouns the correct sense was reliably selected.

Next, we evaluated whether STAn selected a connection between the nouns which was possible according to its world model and plausible for human experts, according to their world knowledge. The triple-noun expressions count as one successful selection only when the entire interpretation is possible. We encountered 84 possible interpretations (or 84%) in our sample.

Obviously, only some of these possible interpretations could be considered the best of all possibilities in the given context. However, STAn also has a chance to rank its preference towards sentence interpretation. In other words, the degree of success in interpreting a noun + noun expression may be subordinated to the correct interpretation of the rest of the sentence. We are interested here in whether the best interpretation, according to the human experts, was found; and the number for that is 69 (69%); again, the triple-noun expressions counted as one unit.

Yet a lower number corresponds to the those correct interpretations that were ranked the highest by STAn:

- 46 of them had the highest rank
- 8 of them had the second-highest rank
- 6 of them had the third-highest rank
- 9 more were lower than the third-highest rank

Therefore, given a choice between TMRs, the selection by STAn and the human experts coincided 66.(6)%, or exactly 2/3 of all cases in the sample.

3) Sources of failure

There are three major reasons for failure, all of which will be removed at the later stages of OST implementation. First, we ran this sample on a limited lexicon, supported by a limited ontology. Both of these are steadily expanded—see [2] for the cost and time estimates for OST acquisition. The Unattested Input Module was not activated in this experiment, thus lowering STAn's robustness.

Second, we marked 9 conspicuous cases of ellipsis, accounting for 60% of all failures to identify the possible connection, and a closer analysis would probably reveal several more cases. As we mentioned before, STAn couldn't handle *group plan* correctly because it lacked the information that it was actually a group insurance plan. Semantic ellipsis remains an underrecognized and underexplored phenomenon, leading researchers to giving up occasionally on interpretations of such expressions—see, for instance, [14] on the meaning of *fast motorway*. OST can interpret the meanings of such expressions by looking at the domain of SPEED—"fast" is in the range of this property—and the conceptually contingent CAR, which is what is fast on the motorway.

For the reconstruction of ellipsis, OST has to use InfoStore, and the choice not to use information in it for this experiment is the third major reason for the failures, especially when the discovery of the best property and TMR ranking are concerned.

These and other possible problems are going to be corrected in the course of further OST implementation.

IV. SUMMARY

We demonstrated, on a limited sample of 100 $N+N^+$ compounds, the principles of analysis and computational interpretation for compositional multi-word expressions in Ontological Semantic Technology. We compared this OST-based methodology with the traditional taxonomy for the meanings of noun compounds and achieved a finer grain size with the former. While the numbers characterize the experiment as reasonably successful, we also indicated the clear directions of improvement in the course of further OST implementation.

ACKNOWLEDGMENT

The authors are grateful to RiverGlass, Inc., for the permission to use their proprietary resources for examples and illustrations.

REFERENCES

[1] S. Nirenburg and V. Raskin, *Ontological Semantics*. Cambridge, MA: MIT Press, 2004.

[2] V. Raskin, J. M. Taylor, and C. F. Hempelman, “Guessing and knowing: Two approaches to semantics in natural language processing,” in *Papers from the Annu. Int. Conf. “Dialogue”*, Moscow, 2010, vol. 9, no. 16, pp. 642-650.

[3] I. A. Sag, T. Baldwin, F. Bond, A. Copestake, and D. Flickinger, “Multiword expressions: A pain in the neck for NLP,” in *Proc. 3rd CICLING*, 2002, pp. 1–15.

[4] T. Finin, *The semantic interpretation of compound nominals*. Ph.D. Dissertation, University of Illinois at Urbana-Champaign, 1980.

[5] R. Girju, D. Moldovan, M. Tatu, and D. Antohe, “On the semantics of noun compounds,” *Computer Speech and Language*, vol. 19, no. 4, pp. 479–496, 2005.

[6] M. Lauer. *Designing Statistical Language Learners: Experiments on Noun Compounds*. Ph.D. Thesis, Macquarie University, Australia, 1995.

[7] D. Ó Séaghdha, and A Copestake, “Co-occurrence contexts for noun compound interpretation.” in *Proc. ACL-2007 Workshop on A Broader Perspective on Multiword Expressions*, Prague, Czech Republic, 2007, pp. 57–64.

[8] M. Lapata, “The disambiguation of nominalizations.” *Computational Linguistics*, vol. 28, no. 3, 2002, pp. 357–388.

[9] J. M. Taylor, V. Raskin, “Fuzzy ontology for natural language,” 29^{th} *Int. Conf. of the North American Fuzzy Information Processing Society,* Toronto, Canada, July 2010.

[10] M. Petrenko, “Lexicon management in ontological semantics,” in *Papers from the Annu. Int. Conf. “Dialogue”*, Moscow, 2010, vol. 9, no. 16, pp. 636-641.

[11] J. M. Taylor, C. F. Hempelmann, and V. Raskin, “On an automatic acquisition toolbox for ontologies and lexicons,” In *Proc. ICAI’10*, Las Vegas, USA, July 2010.

[12] P. Isabelle, Another look at nominal compounds,” in *Proc. COLING-84*, Stanford, CA, USA, 1984.

[13] D. Biber, S. Johansson, G. Leech, and S. Conrad, *Longman Grammar of Spoken and Written English*. Indianapolis, IN: Pearson ESL, 1999.

[14] N. Asher and A. Lascarides, “Metaphor in discourse,” in *Symposium: Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity. Working Notes. AAAI Spring Symposium Series.* J. B. Klavans, B. Boguraev, L. Levin, and J. Pustejovsky, Eds. Stanford, CA: Stanford University, 1995