Treating Unexpected Input in Incremental Semantic Analysis

Marjorie McShane MARGEMC34@GMAIL.COM
Kevin Blissett KEVINPBLISSETT@GMAIL.COM
Irene Nirenburg IRENE_BN@YAHOO.COM
Cognitive Science Department, Rensselaer Polytechnic Institute, Troy, NY 12208 USA

Abstract

Informal language, such as spoken dialog, is full of utterances that fail to conform to neat notions of grammaticality. Presumably, people understand irregular inputs by approximately aligning them with their lexical, syntactic, semantic, ontological, and context-based expectations. If intelligent agents are to function with the linguistic dexterity of people, they must be supplied not only with canonical expectations about grammar, lexicon and the world, but also (a) reasoning capabilities that can derive meaning even when those expectations are incompletely fulfilled, (b) the ability to evaluate their confidence in their interpretations of all inputs, and (c) the ability to determine whether their current level of analysis is actionable or whether additional resources should be expended in pursuit of a more confident analysis. This paper describes how we are enabling language-endowed intelligent agents (LEIAs) to recover from so-called "unexpected inputs" during incremental, deep-semantic analysis. Our strategy: preparing LEIAs to expect the unexpected.

1. Introduction

Descriptions of language, as well as natural language processing (NLP) systems, tend to focus on canonical phenomena – i.e., "grammatical" sentences. Spontaneous natural language, by contrast, often bears little resemblance to idealized models, as illustrated by the following excerpt from the Santa Barbara corpus of spoken language. The speaker is a student of equine science talking about blacksmithing.

> we did a lot of stuff with the -- like we had the, um, ... the burners? you know, and you'd put the -- you'd have -- you started out with the straight .. iron? .. you know? and you'd stick it into the, .. into the, .. you know like, actual blacksmithing (DuBois et al. 2000-2005).[1]

Outside of context, and unsupported by the intonation and pauses of spoken language, this excerpt requires a lot of effort to understand.[2] Presumably, we get the gist by partially matching

[1] Some annotations have been removed for concise presentation.
[2] A reviewer remarked that we are making the task artificially more difficult by providing LEIAs with only a written transcript instead of the speech stream itself, since speech has prosodic features that assist people in extracting meaning. Indeed, prosodic features could be very useful to LEIAs; however, currently unmet is the precondition of automatically extracting and interpreting them within the agent's ontological model. Developing such capabilities is necessary for human-level LEIAs – along with gesture recognition and interpretation, the simulation and

elements of input against the expectations in our mental grammar, lexicon, and ontology (remember, we told you this was about blacksmithing).

For language-endowed intelligent agents (LEIAs), unexpected input includes not only what people would consider non-canonical phenomena but also anything idiosyncratically not covered by the agent's knowledge bases, language processing engines, or reasoning engines. For example, if the agent happens to not know what a grapefruit is, then it will have to resort to unexpected-input processing for the seemingly canonical input *I ate a grapefruit for breakfast*. The results of its recovery are similar to what most people would conclude when faced with the input *Paul ate some cupuacu for breakfast this* morning: it must be some sort of food but it's unclear exactly what kind. (It is a fruit that grows wild in the Amazon rainforest.)

The main claims of this paper are: (1) LEIAs must be able to interpret input that is unexpected with respect to the current state of their knowledge and processors – called, hereafter, *unexpected input*; and (2) many types of unexpected-input processing are possible within the state of the art, as shown by our OntoSem2 system. The body of the paper describes some of the ways in which we are preparing LEIAs to interpret unexpected input. We focus on the following six phenomena: interpreting midstream fragments, preposition swapping, unexpected realization of an internal argument, extra elements of input, unknown words, and known words used in the wrong part of speech. We selected these not because they are inherently more important than other types of unexpected input (for a more comprehensive overview of unexpected input processing by LEIAs, see Nirenburg and McShane, 2016), but because we have new R&D results to report for them. Since this is a considerable list of phenomena, and since we have already described our general approach to cognitively-inspired natural language understanding (NLU) in many places (McShane and Nirenburg, 2012, 2015; McShane et al., 2005, 2008, 2016; Nirenburg and Raskin, 2004), we will constrain the background section to the bare minimum and provide select details in the relevant subsections.

2. Language Understanding by OntoAgents

We pursue deep-semantic natural language processing (NLP) within the agent architecture called OntoAgent (McShane and Nirenburg, 2012). Intelligent agents developed within this architecture have the typical cognitive capabilities of perception, reasoning and action. They can also optionally feature a dynamic physiological simulation, which is useful, for example, when they play the role of virtual patients in medical simulations (Nirenburg et al., 2008; McShane et al. 2012, 2013a,b). We model language understanding as a channel of perception, following the theory of Ontological Semantics (Nirenburg and Raskin, 2004).

The goal of OntoAgent text analysis is to automatically generate fully specified, disambiguated, ontologically-grounded text meaning representations (TMRs) from unconstrained natural language inputs. For example, the TMR for the input *You need to apply pressure to the wound* is as follows. (For reasons of space, we exclude the extensive metadata used by developers – e.g., which word of text is being analyzed, which lexical sense was used to generate each frame, inverse frames unless they include additional properties, etc.)

interpretation of more channels of perception, and much more – but it will not be a near-term focus of our group's research without collaboration with specialists in automatic speech recognition.

REQUEST-ACTION-1			**PRESS-1**		
AGENT	HUMAN-1	*; the speaker*	AGENT	HUMAN-2	*; the interlocutor*
THEME	PRESS-1		THEME	WOUND-INJURY-1	
BENEFICIARY	HUMAN-2	*; the interlocutor*			

This TMR is headed by a numbered instance of the concept REQUEST-ACTION, which is the interpretation of "you need to".[3] The AGENT of this action is the HUMAN speaker and its THEME (what is requested) is an instance of PRESS. The PRESS event instance is further specified, in its own frame, as having the HUMAN interlocutor as its AGENT and an instance of WOUND-INJURY as its THEME. This instance of WOUND-INJURY will be coreferred with an earlier instance (explaining the article *the*) when the full context is analyzed.

The concepts referred to in TMRs are not merely symbols in an upper-case semantics. They are grounded in a 9,000-concept, property-rich ontology developed to support semantically-oriented NLP, script-based simulation, and overall agent reasoning (McShane and Nirenburg, 2012). For example, PRESS is the child of APPLY-FORCE. Among its property-value pairs are case-roles that support lexical disambiguation, including (AGENT ANIMATE), (INSTRUMENT LIMB, DEVICE), (THEME PHYSICAL-OBJECT).

A prerequisite for automatically generating TMRs is our highly specified lexicon. Consider, for example, the first two verbal senses for *address*, shown in Table 1 using a simplified formalism. Syntactically, both senses expect a subject and a direct object in the active voice, filled by $var1 and $var2, respectively.[4] However, in *address*-v1, the meaning of the direct object (^$var2; '^' indicates 'the meaning of') is constrained to a HUMAN or, less commonly, ANIMAL, whereas in *address*-v2 the meaning of the direct object is constrained to an ABSTRACT-OBJECT. The constraints appear in italics because they are virtually available – the analyzer accesses them from the ontology at runtime. 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; by contrast, if the direct object is human, as in *He addressed the audience*, then *address* will be analyzed as SPEECH-ACT.

Table 1. Two verbal senses for the word *address*. The symbol ^ indicates "the meaning of".

address-v1 anno definition "to talk to" example "He addressed the crowd." 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." syn-struc subject $var1 v $var0 directobject $var2 sem-struc CONSIDER AGENT ^$var1 *(sem HUMAN)* THEME ^$var2 *(sem ABSTRACT-OBJECT)*

[3] The obligative modality indicated by *need to* is interpreted as a request for action when addressed to the interlocutor.

[4] Variables are written, by convention, as $var followed by a distinguishing number. Variables permit the language analyzer to map textual content from the input to elements of the syn-struc, and then link each syn-struc element with its semantic realization in the sem-struc.

These examples highlight several aspects of our lexicon. First, it supports the combined syntactic and semantic analysis of texts. Second, the metalanguage for describing meaning in the sem-strucs is the same one used in the ontology. Third, fixed and variable constructions of any complexity are readily supported (McShane et al., 2015). Finally, the sem-strucs—and, often, the associated syn-strucs—from the lexicon for one language can be ported into the lexicon of another language with minimal modification, which greatly enhances the multilingual applicability of the OntoAgent suite of resources (McShane et al., 2005).

In this paper, we will not describe in detail the new implementation of Ontological Semantics that we call OntoSem2 (for a comparison with the previous engine, OntoSem, see McShane and Nirenburg, 2016). The key features of this implementation are: (1) it processes inputs incrementally, rather than as full sentences, which is psychologically more plausible and offers important practical advantages, such as allowing agents to begin acting on an input midstream; (2) it utilizes syntactic, semantic and reference heuristics jointly, rather than in a pipeline architecture; (3) it exploits select results of the Stanford CoreNLP toolset (Manning et al., 2014); and (4) it attempts not only basic semantic analysis – a difficult feat in and of itself – but also the analysis of many types of particularly difficult phenomena, such as ellipsis, broad referring expressions, and learning unknown words and concepts (McShane and Babkin, 2016a,b).

3. Recovering from Unexpected Input

Our strategy for recovering from unexpected input is to anticipate and explicitly prepare for as many eventualities as possible, rendering them functionally *almost expected*. Each of the six subsections below discusses implemented methods for recovering from the types of unexpected input listed in the introduction.

3.1 Interpreting Midstream Fragments

It might seem surprising to consider the interpretation of midstream sentence fragments a type of unexpected input since analyzing fragments is the very definition of incremental parsing. However, treating fragments requires two types of processing not typical for NLP:

(1) Reasoning over partial information: i.e., determining which candidate word senses and dependencies can confidently be excluded given the current state of input and which ones are still available. For example, English verbs do not normally take both a direct object and a clausal complement, so if one of them is confidently attested in a fragmentary input, then all verb senses requiring the other can be excluded.
(2) Carrying ambiguity forward, which is avoided by many pipeline architectures to avoid the potential for exponential growth of candidate analyses stage after stage.

Consider the incremental semantic analysis of the input *Audrey killed the motor*, presented with only select details for reasons of space. The first word of input is *Audrey.* The onomasticon contains only one sense of this string, so the nascent TMR is

HUMAN-1

GENDER	FEMALE
HAS-PERSONAL-NAME	'Audrey'

The next word is *killed*, so the combination *Audrey killed* is analyzed. The lexicon has 5 senses of *kill*, but only three of them permit a HUMAN to fill the subject slot: 1) 'cause to die': *John killed the robber*; 2) 'cause to cease operating': *John killed the engine*; and 3) 'veto': *The committee killed the bill*. The other two senses can be excluded outright since one requires the subject to be an event (*The disease killed him*) and the other requires it to be a non-human object that serves as an instrument (*The bullet killed him*).[5]

The next word of input is *the*. The LEIA does not launch a new round of semantic analysis for the fragment *Audrey killed the* because no useful information can be gleaned from function words without their heads.

The next and final stage of analysis is launched on the entire sentence *Audrey killed the motor*. Each of our three still-viable senses of *kill* includes semantic constraints on the direct object: for sense 1 it must be an ANIMAL, for sense 2, an ENGINE, and for sense 3, a BILL-LEGISLATIVE. Since 'motor' maps to the concept ENGINE, sense 2 – 'cause to cease to operate' – is selected and the final TMR for *Audrey killed the motor* is as follows:

ASPECT-1		**OPERATE-DEVICE-1**		**HUMAN-1**	
PHASE	END	AGENT	HUMAN-1	GENDER	FEMALE
SCOPE	OPERATE-DEVICE-1	THEME	MOTOR-1	HAS-NAME	'Audrey'
		TIME	< find-anchor-time[6]		

We selected to illustrate incremental processing using the simplest of examples for purposes of clarity. In reality, most sentences involve much more midstream ambiguity (i.e., many more candidate analyses), and it is not atypical for the LEIA to be unable to fully resolve the ambiguity based on ontological and lexical constraints alone – other contextual heuristics can be needed, which is the topic of a different paper in progress. The point here is that fragmentary inputs can be classified as "unexpected input" because not all of an argument-taking word's needs are immediately fulfilled, making it necessary for the LEIA to reason over partial information. We have implemented this reasoning by making the LEIA disambiguate only to the degree that a person given the same fragmentary input could disambiguate.

3.2 Preposition Swapping

Prepositions are a common source of performance error by non-native speakers and native speakers alike. For example, the subtitles for the Finnish TV series *Easy Living* are of very high quality overall but include some unusual preposition selections. Considering that English is the current *lingua franca*, with many speakers having non-native fluency, it is of high priority for LEIAs to accommodate this type of close-but-not-perfect input. The following examples from the COCA corpus (Davies, 2008) illustrate this phenomenon. They all use the expression *translate **to*** in place of the canonical *translate **into*** in reference to translating languages:[7] *And you can feel this tension with every sentence that you say, and this tension can not [sic] be **translated to** any*

[5] For simplicity's sake, this example assumes the active voice.

[6] This filler for the TIME slot is a call to a procedural semantic routine that attempts to ground the moment of speech in real time, and then analyzes the past tense as "before that actual time".

[7] There are several non-literal uses of *translate to* (not *into*) that expect different types of arguments semantically, as shown by the following examples from COCA: *That relationship **translated to** a better learning environment; …but the controls are just like a video game's—only **translated to** a physical toy; … advancements in diagnosis and care as well as treatment in humans, is **translated to** pets — specifically canines…*

other language; NPR also reported the context of the slip-up, ***translated to*** *English:* "*If each one of us does not amass riches only for oneself, but half for the service of others, …*"

Clearly it is not that case that any preposition can be swapped for any other in any context. We have formulated the following three conditions, all of which must be met before a LEIA includes a "preposition swapping" analysis among its candidate analyses for the input.

1. The lexicon must contain a fixed expression (i.e., an idiom or construction) that matches the input lexico-syntactically except for the preposition choice. So we are not talking about free combinations of prepositions and their complements.
2. All of the semantic constraints for that fixed expression must fulfilled. In our example of *translate X into [to] Y*, X must be a language or text and Y must be a language. These constraints are specified in the lexical sense for canonical *translate X into Y.*
3. The preposition pair (here: *into/to*) belongs to a list of preposition pairs that we have found to be, or hypothesize to be, subject to swapping. These pairs either contain prepositions with similar literal meanings (*in/into, into/to, from/out of, by/with*, etc.) or they contain at least one preposition that is extremely semantically underspecified, such as *of.*

As evidence of the need to recover from unexpected preposition use, consider the following examples from the COCA corpus, which contain unexpected prepositions as judged against the expectations recorded in our lexicon. The canonical, lexically listed prepositions are indicated in { } after the unexpected, attested ones: *... if we pull back the aid, they will no longer abide with {by} the treaty with Israel; No, I can not [sic] be absolved from {of} my blame; Changes indicated by the validation panel and field test were incorporated in {into} the instrument development; Practice began this week in anticipation for {of} the season opener at Virginia on Aug. 30.*

The question then becomes, if we find an attested example of an unexpected prepositional use, should it be recorded in the lexicon and treated ever after as canonical? The answer is "No" for three reasons. First, when we generate language we do not want to generate the less-preferred version. Second, our recovery procedure should be sufficient to detect the swapping dynamically and support a successful analysis, making static lexical recording unnecessary. Third, resorting to a recovery procedure models the additional cognitive load of processing unexpected input; this will result in a scoring penalty, meaning that successful analyses that are carried out *without* unexpected input processing will be preferred, as they should be. All that being said, the LEIA's history of language analyses could be consulted when judging confidence in unexpected input processing. For example, if the LEIA found that particular preposition swap multiple times in its repository of past analyses, it could reduce the penalty for that swap to a fraction of the norm.

3.3 Unexpected Realization of an Internal Argument

One type of unexpected input attested in a recent evaluation (McShane et al., 2016) occurs when an input could have been correctly semantically analyzed using a verb sense available in our lexicon but the internal arguments were syntactically realized in an unexpected way. For example, at the time of evaluation, the lexicon covered the structure *He begged me to come* but not *He begged that I come.*

Recorded:	subject	+	V +	direct object	+	infinitive
Attested:	subject	+	V +	[$_{\text{COMPL}}$ subject of clausal compl.	+	VP of clausal compl.]

Given a syntactic mismatch like this, the question is: Is the recorded sense a syntactic variant of the attested sense, or is the attested sense something differently entirely? The LEIA attempts to determine this by first aligning the syntactic constituents using recorded, expectation-driven matching strategies (we don't expect any magic to happen here – or elsewhere, for that matter), and then determining whether the semantic constraints on the arguments are met. For example, the subject and direct object of *beg* in our recorded sense must both be ANIMALs, but there are no constraints on the meaning of the infinitival complement. This means that, in the attested input, the subject of the main clause and the subject of the embedded clause must both be ANIMALs as well.

Our current implementation covers all verbs with the abovementioned syntactic expectations. We expect further corpus study to reveal other canonical syntactic pairings that will be useful to record as expected correlations.

3.4 Extra Elements of Input

In some cases, after all of the valencies of argument-taking words have been accounted for, extra words remain in the input. This category of unexpected input is well illustrated by our earlier Santa Barbara Corpus example. Consider a slightly edited version of an excerpt from that example: *You'd stick it into the into, actually, into the actual oven.* This includes the repetition of an incomplete constituent (*into the into*) and a parenthetical expression (*actually*) that intervene in what is in fact a quite simple sentence: *You'd stick it into the actual oven*. Although people find it easy to cut through this clutter – particularly in the spoken language, when intonation and pauses help – not so for LEIAs. The goal is to detect and strip away the superfluous elements, leaving only the core ones behind. We do this by stripping a list of interjections as well as repetitions of unigrams, bigrams, etc., up to 5-grams. So, the LEIA strips strings as indicated by the strikethroughs: *When the, ~~uh, uh,~~ ~~when the~~ ship was beginning to move*; *~~The,~~ the dog came from Puppy Jake Foundation."*[8]

Stripping superfluous strings builds upon two past (and ongoing) threads of work that seek to focus the LEIA's attention on the core elements of input: dynamic tree trimming, which we have applied successfully to the analysis of verb phrase ellipsis (McShane and Babkin, 2016a); and goal-oriented methods of achieving actionable – in contrast to necessarily comprehensive – interpretations of input (McShane and Nirenburg, 2015).

3.5 Unknown Words

The OS lexicon contains about 30,000 word senses, making it substantial but far from comprehensive. This means that LEIAs must be able to process both unknown words and unknown senses of known words – eventualities that we treat in turn in this and the following subsections.[9]

Processing unknown words begins by using the part-of-speech tag provided by Stanford's CoreNLP toolset (Manning et al., 2014), which we assume to be correct due it high precision overall. We have created generic default lexicon entries for the main parts of speech in a variety

[8] The first is an excerpt from a COCA corpus example and the second is a full COCA example.
[9] For work on learning new words in a prior implementation of Ontological Semantics, see Nirenburg et al. (2007).

of syntactic environments that serve as the anchor for reasoning about semantics. We illustrate unknown word analysis using three simple examples of an unknown noun, verb, and adjective, in turn.

Unknown Noun. Syntactically, simple nouns take no arguments. Semantically, they can refer to an object, event, or property; so the LEIA generates three candidate senses for each unknown noun that allow for each of these semantic eventualities.[10] Each of these candidate senses is then evaluated within the rest of the context, not only to choose the best one but, in many cases, to narrow down the interpretation to a more specific ontological type.

Consider the example *A brooze was sleeping in the park*. The lexicon contains senses for all of these words except for *brooze*. Since *brooze* is selected as the subject by the verb *sleep*, and since the verb *sleep* maps to the concept SLEEP, the LEIA can use its knowledge about SLEEP to narrow down what a *brooze* must be. Specifically, the subject of *sleep* maps to the EXPERIENCER of SLEEP, and the EXPERIENCER of SLEEP is ontologically defined as being an ANIMAL. This tells the LEIA two things. First, of its original three candidate senses of *brooze* – which mapped to OBJECT, EVENT, and PROPERTY – the OBJECT sense is the correct one. Second, the OBJECT analysis can be further constrained to ANIMAL, since only ANIMALs can experience SLEEP. So the final TMR for this example (as before, minus metadata and inverses) is:

ASPECT-1		**SLEEP-1**	
PHASE	CONTINUE	EXPERIENCER	ANIMAL-1
SCOPE	SLEEP-1	LOCATION	PARK-1
		TIME	< find-anchor-time

The metadata carries a trace that unknown-word recovery was carried out, should the LEIA decide to pursue a more fine-grained analysis of this word through learning by reading (Nirenburg et al., 2007) or by interacting with a human collaborator (McShane and Nirenburg, 2015).

Our second example – *Jane was eating kuzdra with a knife* – shows what happens when a case-role slot has a *set* of semantic constraints rather than just one. The verb *eat* has several senses in our lexicon, all but one of which cover idiomatic constructions that are rejected on lexico-syntactic grounds (e.g., *eat away at*); so the LEIA can immediately narrow the choice space to the main sense of *eat*, which is optionally transitive and maps to an INGEST event whose case-roles are AGENT and THEME. The THEME of INGEST is ontologically specified as the disjunctive set DRUG or INGESTIBLE. Since both drugs and ingestibles are of the type OBJECT, the LEIA selects the OBJECT candidate sense for analyzing the unknown word *kuzdra* (rejecting the EVENT and PROPERTY senses). As before, it uses ontological constraints to narrow down the interpretation but, this time, both members of the disjunctive set must be permitted. Formally, this is done with the typical set notation used in TMRs, making the final TMR for this input as follows:

[10] This is functionally equivalent to creating a single sense with a semantic disjunction, which was technically more cumbersome.

ASPECT-1

PHASE	CONTINUE
SCOPE	INGEST-1

INGEST-1

AGENT	HUMAN-1
INSTRUMENT	KNIFE-1
THEME	SET-1
SCOPE-OF	ASPECT-1
TIME	< find-anchor-time

HUMAN-1

AGENT-OF	INGEST-1
HAS-NAME	Jane

SET-1

THEME-OF	INGEST-1
SET-MEMBER-TYPE	DRUG INGESTIBLE
SET-TYPE	disjunctive

Unknown Adjective. Syntactically, adjectives modify a noun. Semantically, they map to a PROPERTY, and the noun they modify fills the DOMAIN slot of that property. The RANGE, however, depends on the actual meaning of the adjective. For example, the subject noun phrase in the sentence *A wugly police officer was driving a red car* includes the unknown adjective *wugly.* The LEIA's analysis of *wugly police officer* is (POLICE-OFFICER-1 (DOMAIN-OF PROPERTY-1)), with PROPERTY-1 representing an instance of the most underspecified ontological property. If asked to, the LEIA can generate the entire set of properties for which POLICE-OFFICER is a semantically acceptable filler for DOMAIN. Although this would be far smaller than the full set of properties in the ontology, in most cases it will be too large to be of much more utility than the generic PROPERTY. The point, however, is that the LEIA's understanding of unknown adjectives mirrors that of a person: we, too, don't know what *wugly* is beyond the fact that it is a property that can be used to describe police officers.

Unknown Verb. Syntactically, verbs can take various numbers and types of arguments and complements, and these can be realized semantically in various ways. The LEIA reasons about the syntax-to-semantics linking using an inventory of default expectations. For example, transitive verbs most often realize the subject as the AGENT and the direct object as the THEME; however, if the subject cannot, semantically, be an AGENT (e.g., if it is an EVENT or INANIMATE-OBJECT), then the most likely candidates are THEME or INSTRUMENT, depending on (a) the meaning of the entity and (b) which other case-roles have already been spoken for by the input. For example, given *The truck bloophed the tree*, *the truck* cannot be the AGENT since it is inanimate. So the LEIA evaluates whether it can be the THEME: it cannot, because the THEME case-role is already taken. So it selects the next-in-line case-role for inanimates: INSTRUMENT.

Now consider the example *John kuzdered the woman for the argument*. Syntactically, the LEIA analyzes this as a transitive verb with a free PP adjunct. Semantically, it maps *kuzdred* to the most generic EVENT, filling its AGENT slot with the interpretation of *John* (HUMAN-1 (HAS-PERSONAL-NAME John)), filling its THEME slot with the interpretation of *the woman* (HUMAN-2 (GENDER FEMALE) (AGE (> 15))), and interpreting the free PP adjunct *for the argument* as an ARGUE event that is the PURPOSE of 'kuzdred'. Of course, it is impossible for people, no less LEIAs, to know whether PURPOSE is the best case-role selection for representing the meaning of the PP adjunct. In fact, regular sentences (those lacking unknown words) with PP adjuncts often result in residual ambiguity due to the extreme challenges of disambiguating highly polysemous prepositions. However, the quite separate problem of analyzing free PP adjuncts should not detract from point of this example: that default correlations between syntactic arguments and semantic roles can result in a reasonable, albeit underspecified, interpretation for inputs with unknown verbs. As in previous examples, the LEIA could also attempt to narrow down the interpretation of the underspecified EVENT based on unidirectional case-role constraints: here, *kuzder* can refer to only those events that permit a HUMAN as the AGENT and a HUMAN as the THEME – a large subset, to be sure, but much smaller than the entire inventory of events in the ontology.

3.6 An Available Lexical Sense is of the Wrong Part of Speech

It is not unusual for the lexicon to contain a needed string but in the wrong part of speech: e.g., a text might contain the verb *heat* (*A large radiator was heating the room*) whereas the lexicon contains only the noun *heat*. The first thing to say about such situations is that there are many possible eventualities. For example, the lexicon might contain exactly one nominal sense which, luckily, is semantically related to the needed verbal sense; the lexicon might contain multiple nominal senses, one of which is related to the needed verbal sense; or the lexicon might contain one or more nominal senses, none of which is related to the needed verbal sense. Similar sets of eventualities can obtain when a known nouns seeds the analysis of an unknown verb. Our basic approach to these cases is as follows. For exposition (and to avoid endless singular/plural options), we assume that there is just one recorded sense of each string; if there is more than one, the process iterates over all possibilities and generates multiple options.

Let us work through the abovementioned example, *A large radiator was heating the room*, in which a recorded nominal sense must be used to analyze an unknown verb. This example is particularly interesting because the noun in question, *heat*, maps not to an OBJECT or EVENT, but to an ontological PROPERTY: TEMPERATURE. For reasons described in McShane et al. (2008), we treat CHANGE-EVENTs specially, since their meaning best captured by comparing the value of the given property in the PRECONDITION and EFFECT of the CHANGE-EVENT. For example, *accelerate* indicates an increase in the value of SPEED, *shrink* indicates a decrease in the value of SIZE*,* and *fatten* indicates an increase in the value of WEIGHT. The TMR for the example *A large radiator was heating the room* is shown below.

ASPECT-1	
PHASE	CONTINUE
SCOPE	CHANGE-EVENT-1

CHANGE-EVENT-1	
THEME	ROOM-1
INSTRUMENT	HEATER-1
SCOPE-OF	ASPECT-1
PRECONDITION	TEMPERATURE-1
EFFECT	TEMPERATURE-2
TIME	< find-anchor-time

TEMPERATURE-1	
DOMAIN	ROOM-1
RANGE	< TEMPERATURE-2.RANGE

TEMPERATURE-2	
DOMAIN	ROOM-1
RANGE	> TEMPERATURE-1.RANGE

HEATER-1	
SIZE	.7
INSTRUMENT-OF	CHANGE-EVENT-1

Consider how much agent reasoning required to arrive at this analysis. (1) The LEIA finds the noun *heat* described as (TEMPERATURE (RANGE (> .8))). (2) It recognizes this as a CHANGE-EVENT situation: a noun with a property-based description is being used to analyze an unknown verb. (3) It instantiates a CHANGE-EVENT along with PRECONDITION and EFFECT slots. (4) It hypothesizes the direction of change (i.e., the comparison between the range of TEMPERATURE in the PRECONDITION and EFFECT) based on the RANGE of the property in the nominal sense: if the nominal sense has a high value (like our .7), then it assumes that the direction of change is increase, whereas if the nominal sense has a low value, then it assumes that the direction of change is decrease. (5) It interprets the THEME of the CHANGE-EVENT (ROOM-1) as the DOMAIN of the TEMPERATURE frames. (6) It deals with other semantic analysis needs, such as lexical disambiguation of the theme and the analysis of tense and aspect

4. Comparison with Others

Full-fledged computational semantics has not been pursued by the mainstream natural language processing community for over twenty years. The programs of work that most closely resonate with ours peaked before the so-called statistical revolution in NLP: Schank's Conceptual Dependency Theory (Schank, 1972) and Wilks' Preference Semantics (Wilks, 1985). As concerns *incremental* semantic analysis, it was attempted already in the 1980s, e.g., by Mellish (1985) and Hirst (1988). Both Mellish and Hirst draw analogies between incremental semantic analysis and vision research, with Hirst calling his approach *Polaroid Words*. More recent work on incrementality is represented by Jerry Ball's incremental syntactic parser (Ball et al., 2014) and Ruth Kempson's (2000) Dynamic Syntax. Although the latter is a purely linguistic (rather than computational linguistic) theory, it was used by Purver at al. (2011) as the theoretical substrate for an incremental parser.

A newcomer to cognitively-inspired language processing is a the program of work reported in Lindes and Laird (2016). It is gratifying to see another research group addressing the issues – and, in general terms, embracing the approaches – that Ontological Semantics has been long pursuing. We second Lindes and Laird's claim that all of the following ten features are crucial for human-level language processing: the processing must be *incremental*, *integrated, eclectic, useful*, and carried out in *real time*; it must be *compositional, hierarchical*, and *grounded*; it must compute *context-dependent meaning* and include *repair-based processing*. It is noteworthy that Lindes and Laird echo our use of "actionable" interpretations (McShane and Nirenburg, 2015) as the goal of language-endowed intelligent agents.

5. Discussion

In developing agent-based language understanding capabilities, one can focus on domain-independent strategies, domain-specific strategies for which agents have extensive knowledge and reasoning support, or strategies that fall somewhere in between. This work contributes to the domain-independent thread – as has our recent work on processing various other types of difficult linguistic phenomena, such as verb phrase ellipsis (McShane and Babkin, 2016a) and broad referring expressions (McShane and Babkin, 2016b).

There are both scientific and tactical reasons for pursuing domain-independent strategies in addition to domain-specific ones, even if one's overall goal is configuring human-level intelligent agents that will necessarily have extensive domain knowledge. On the scientific side, strategies like the ones described above actually *are* domain-independent and reflect domain-independent reasoning, so casting them as such is correct from the perspective of human-inspired cognitive modeling. On the tactical side, they can be launched over any corpus and contribute to the work of mainstream NLP, for which large corpora – and the non-reliance on knowledge support – are the object of interest. So, even though our LEIAs cannot, today, generate completely correct semantic interpretations of every input (additional well-understood types of knowledge engineering are needed to increase system coverage), they can analyze many inputs in the general domain with high confidence, and those analyses could be leveraged in interesting ways, we think, by the statistical NLP community.

By contrast, recovering from unexpected input in narrow domains will require supplementing these domain-independent strategies with a battery of knowledge-based, reasoning-heavy processes that leverage the LEIA's knowledge about its own plans and goals, those of its interlocutor, its understanding of the situation, etc. We leave the description of our approach to these kinds of recovery for a future paper, but interested readers can find relevant past work in our descriptions of the Maryland Virtual Patient application (McShane et al., 2012, 2013a, 2013b).

Of the many additional issues we could address in this discussion, we focus on three that were noted by reviewers: coverage of phenomena, the state of implementation of the system, and evaluation.

Coverage of phenomena. The six types of unexpected input described here represent a subset of the unexpected-input phenomena that LEIAs will need to treat. For example, multi-party interactions feature overlapping utterances, interruptions, and people finishing each other's sentences; noisy-channel inputs can render some fragments uninterpretable/untranscribable; and transcripts produced by speech recognition systems are subject to error. (The latter, of course, means that word recognition should ideally be approached by comparing candidate transcriptions with the LEIA's context-based expectations about which words and phrases are most likely to occur.) Moreover, not only are there many classes of unexpected input, their realizations can be combined so densely in naturalistic discourse that the task must change from translating non-canonical inputs into their canonical counterparts, to trying to compose islands of meaning from a sea of uninterpretables. McShane and Nirenburg (2015) describes the latter in the context of pursuing *actionable* language analyses, which can be done by agents in a goal-directed manner. Our lab's plans for the summer and fall of 2017 include the development of a larger, example-supported classification of unexpected input through the analysis of speech corpora. We will then develop programs to convert the unexpected phenomena into their expected-input counterparts. We intend to report on this at next year's ACS conference.

The state of implementation. All six of the recovery phenomena reported in this paper were implemented and tested within the overall OntoSem2 system. The test suite is as follows (see related discussion with respect to evaluation).

- **Preposition swapping**: "They will not abide *with* the treaty with Israel." (by/with); "He offset the lost time *with* working more." (by/with); "This absolved him *from* the obligation." (of/from); "They booted him out *from* the club." (of/from); "They incorporated faith *in* the conversations." (into/in); "A woman was harassed by a convict employed to enter data *in* a database." (into/in); "Practice began in anticipation *for* the game." (of/for)
- **Stripping repetitions and interjections**: "*When the, uh, uh,* when the ship was beginning to move → When the ship was beginning to move." "*The,* the dog came from Puppy Jake Foundation. → The dog came from the Puppy Jake Foundation."
- **The lexicon includes the verb but with different syntactic expectations**: "He begged me to come. → He begged that I come." "He backed the car up. → He backed up the car."[11]

[11] The latter example generalizes to all of the verbs in the lexicon with the syntactic structure "subject + verb + {prepositional particle and direct object, in either order.} We record on word order and derive the other (though we

- **The lexicon contains the string as a verb but not as a noun**: "He won the *award*." "I had a good *cry*." "Get a good *grip*."
- **The lexicon contains the string as a noun but not as a verb**: "The *drizzle* stopped." "My *jog* was great!" "A large radiator was *heating* the room."
- **Unknown adjective**: "The *wugly* man slept."
- **Unknown noun**: "The *wug* slept."
- **Unknown verb**: "I *wug* the house on the street." "I *wugged* the bread." "I *wugged*."

OntoSem2's recovery from these unexpected inputs worked as expected and the system generated at least one candidate TMR for each.

Describing the overall capabilities of OntoSem2 to a useful level of detail would require much more space, so a thumbnail sketch will have to suffice. OntoSem2 can generate TMRs for inputs with the following features: (1) they syntactically conform to a mid-sized grammar of typical syntactic structures, or they can be made to conform to this grammar by unexpected-input processing of the types described here; (2) either all of the lexemes are attested in the OntoSem lexicon in the expected parts of speech, or there is a single unknown one that the system can recover from using the procedures described above. Most words in any input are polysemous, and our 30,000-sense lexicon does not shy away from recording polysemy. This results in multiple candidate TMRs for most inputs. The LEIA automatically scores each TMR candidate based on an inventory of heuristics.[12] The score approximates the agent's confidence in that interpretation as the correct one for the input. The OntoSem2 implementation is currently ongoing. Our team is small, and the stack of phenomenon-treatment specifications ready to be implemented is, not surprisingly, still large.

Evaluation. The above logistics-related statements offer part of the explanation for why this paper does not include a formal evaluation: formal evaluations are expensive and they compete for developer time. But the issue of evaluation raises an even more important question: What added value would a formal evaluation have for the content and goals of *this particular paper*?

In the spirit of the times, people expect numerical evaluations of all work in NLP. However, practically all modern-day NLP systems are statistical rather than knowledge-based, and widely-accepted evaluation practices reflect this orientation. For example, Resnik and Lin's (2010) book chapter, entitled *Evaluation of NLP Systems*, does not even consider trying to evaluate the kinds of scientific goals we pursue. They write, "…Such scientific criteria have fallen out of mainstream computational linguistics almost entirely in recent years in favor of a focus on practical applications, and we will not consider them further here." (p. 271)

In order to *usefully* evaluate a psychologically-inspired, knowledge-based, cognitive system, the evaluation must be phenomenon-specific, carefully designed, and rigorously argued for – otherwise readers will not be convinced that it can count as evaluation. Moreover, the most scientifically interesting aspects of an evaluation involve error analysis and correction, the status

could, alternatively, have recorded duplicate lexical senses for all such cases as well; these knowledge representation methodologies will produce the same processing results).

[12] For a list of heuristics in the original OntoSem implementation, which is largely followed here as well, see McShane, Nirenburg and Beale (2016).

of which can only be understood on the basis of no small number of system details. In published reports, all of this takes up space, making it unsuitable for a conference paper.

In addition, if we *were* to provide a formal evaluation of what we have presented here, what should it look like and why? Before trying to answer that directly, let us start with some facts that we believe do not require evidence beyond what we have presented: (1) The phenomena we describe exist in naturalistic speech. (2) Syntactic parsers, including the one we use, do better on canonical syntactic inputs than on non-canonical ones. (3) OntoSem2's main analysis routine requires a valid syntactic parse, which helps the analyzer to carry out lexical disambiguation and the establishment of the semantic dependency structure. If syntactic analysis fails, so, too, does semantic analysis – at least currently (we are working toward removing this brittleness). (4) Different corpora representing different language genres will have very different amounts and manifestations of unexpected input, so the results of any such evaluation would not be generalizable. (5) The utility of new linguistic descriptions and associated algorithms is not inexorably linked to the overall quality of a particular agent system at a particular time; i.e., overall OntoSem2 evaluation is relatively inconsequential to this report. If readers are to use the information presented here, it will most likely be by using these descriptions to facilitate implementations in their own systems.

One could argue – and many do – that the only worthwhile evaluation involves showing an improvement in the agent's behavior in an end application due specifically to the newly reported functionalities. We agree that this is the gold standard. But the question then becomes, whence the test application? There are at least three possibilities. First, one can develop a fully deployed, useful system and add functionalities to it over time. For the MVP system – our most realistic candidate – this would cost on the order of $20M to just get the ball rolling. Another option is to use a smaller prototype system that has the potential to grow into a useful end application. We are, in fact, working toward this but do not expect to have reportable convergence of the agent's full inventory of perception-reasoning-action capabilities in the very short term. Under this constraint it could take years for us to be able to report tangible progress on the knowledge end, which would be unfortunate. Finally, one can invent a toy application solely for evaluating a particular new functionality; but not only would this be very time-consuming and expensive, toy applications tend to elicit even more criticism than missing evaluations.

Over the years our group has invented novel evaluation strategies for various aspects of our language processing systems, such as lexical disambiguation (McShane, Nirenburg & Beale, 2016), the understanding of multi-word expressions (McShane, Nirenburg & Beale, 2016), and the resolution of verb phrase ellipsis (McShane & Babkin, 2016a). Each evaluation has taken months, and the knowledge-based analysis of errors has occupied a large portion of the associated journal articles. We mention this in closing to emphasize that we are in favor of evaluation when we can see a clear reason for it but, with respect to this particular knowledge-oriented contribution in this particular venue, we believe that the candle would not have been worth the flame.

Acknowledgments. This research was supported in part by Grant N00014-16-1-2118 from the U.S. Office of Naval Research and by the Cognitive and Immersive Systems Laboratory collaboration between Rensselaer Polytechnic Institute and IMB Research. Any opinions or findings expressed in this material are those of the authors and do not necessarily reflect the views of the Office of Naval Research, Rensselaer Polytechnic Institute, or IBM Research.

References

Ball, J., with Freiman, M., Rodgers, S., & Ball, A. (2014). Double-R Grammar: A Computational Cognitive Grammar of English. Last Updated 27 August 2014. Available at http://www.doublertheory.com/double-r-grammar/index.html.

Davies, M. 2008. The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at http://corpus.byu.edu/coca.

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., & Martey, N. (2000-2005). *Santa Barbara Corpus of Spoken American English, Parts 1-4*. Philadelphia: Linguistic Data Consortium.

Hirst, G. (1988). Resolving lexical ambiguity computationally with spreading activation and Polaroid Words. In: Small, Steven L.; Cottrell, Garrison W.; and Tanenhaus, Michael K. (editors), *Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann Publishers, 1988. 73–107.

Kempson, R., Meyer-Viol, W., & Gabbay, D. (2000). *Dynamic Syntax*. Wiley-Blackwell.

Lindes, P., & Laird, J. E. (2016). Toward integrating cognitive linguistics and cognitive language processing. *Proceedings of the 14th International Conference on Cognitive Modeling (ICCM).* University Park, Pennsylvania.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations,* pp. 55–60. Stroudsburg, PA: The Association for Computational Linguistics.

McShane, M., & Babkin, P. (2016a). Detection and resolution of verb phrase ellipsis. *Linguistic Issues in Language Technology* 13(1): 1-34.

McShane, M., & Babkin, P. (2016b). Resolving difficult referring expressions. *Advances in Cognitive Systems*, Vol. 4, pp. 247-263.

McShane, M., Beale, S., Nirenburg, S., Jarrell, B., & Fantry, G. (2012). Inconsistency as a diagnostic tool in a society of intelligent agents. *Artificial Intelligence in Medicine*, 55(3): 137-48.

McShane, M., & Nirenburg, S. (2012). A knowledge representation language for natural language processing, simulation and reasoning. *International Journal of Semantic Computing,* 6(1), 3-23.

McShane, M., & Nirenburg, S. (2015). Decision-making during language understanding by intelligent agents. *Artificial General Intelligence*, Volume 9205 of the series Lecture Notes in Computer Science, pp. 310–319.

McShane, M,. & Nirenburg, S. (2016). Decisions for semantic analysis in cognitive systems. *Proceedings of Advances in Cognitive Systems 2016.*

McShane, M., Nirenburg, S. & Beale, S. (2005). An NLP lexicon as a largely language independent resource. *Machine Translation*, 19(2): 139–173.

McShane, M., Nirenburg, S. & Beale, S. (2008). Ontology, lexicon and fact repository as leveraged to interpret events of change. In Huang, C, Calzolari, N., Gangemi, A., Lenci, A., Oltramari, A. & Prevot, L. (Eds.), *Ontology and the Lexicon: A Natural Language Processing Perspective*, pp. 98-121. Cambridge University Press.

McShane, M., Nirenburg, S., & Beale, S. (2015). The Ontological Semantic treatment of multiword expressions. *Lingvisticæ* Investigationes, 38(1): 73-110. John Benjamins Publishing Company.

McShane, M., Nirenburg, S. & Beale, S. (2016). Language understanding with Ontological Semantics. *Proceedings of Advances in Cognitive Systems 2016*.

McShane, M., Nirenburg, S., Beale, S., Jarrell, B., Fantry, G., & Mallott, D. (2013a). Mind-, body- and emotion-reading. *Proceedings of IACAP 2013 (International Association for Computing and Philosophy)*. University of Maryland, College Park.

McShane, M., Nirenburg, S. & Jarrell, B. (2013b). Modeling decision-making biases. *Biologically-Inspired Cognitive Architectures* 3: 39–50.

Mellish, C. (1985). Incremental semantic interpretation in a modular parsing system. In K. Sparck-Jones and Y. Wilks, (Eds.), *Automatic Natural Language Parsing*. Ellis Horwood Limited/Halstead Press, a division of John Wiley & Sons: New York, Brisbane, Chichester, Ontario, pp. 148-155.

Nirenburg, S. & McShane, M. (2016). Slashing metaphor with Occam's Razor. *Proceedings of the Fourth Annual Conference on Advances in Cognitive Systems*.

Nirenburg, S., McShane, M. & Beale, S. (2008). A simulated physiological/cognitive "double agent". In Beal, J., Bello, P., Cassimatis, N., Coen, M. & Winston, P. (Eds.), *Papers from the AAAI Fall Symposium, Naturally Inspired Cognitive Architectures*. AAAI technical report FS-08-06, Menlo Park, CA: AAAI Press.

Nirenburg, S., Oates, T. & English, J. (2007). Learning by reading by learning to read. *Proceedings of the International Conference on Semantic Computing*. San Jose, August.

Nirenburg, S. & Raskin, V. (2004). *Ontological Semantics*. Cambridge, MA: The MIT Press.

Purver, M., Eshghi, A., & Hough, J. (2011). Incremental semantic construction in a dialogue system. *Proceedings of the 9th International Conference on Computational Semantics* (IWCS), Oxford.

Resnik, P., & Lin, J. 2010. Evaluation of NLP systems. In Clark, A., Fox, C., & Lappin, S. (Eds.), *Handbook of Computational Linguistics and Natural Language Processing*. Wiley Blackwell.

Schank, R. (1972). Conceptual Dependency: A theory of natural language understanding. *Cognitive Psychology* 3: 532–631.

Wilks, Y. (1985). Text structures and knowledge structures. *Quaderni di Semantica, 6*, 335–344.