Advances in Cognitive Systems 4 (2016) Submitted 4/2016; published 6/2016

Slashing Metaphor with Occam's Razor

Sergei Nirenburg ZAVEDOMO@GMAIL.COM
Marjorie McShane MARGEMC34@GMAIL.COM
Cognitive Science Department, Rensselaer Polytechnic Institute, Troy, NY, 12180

Abstract

We argue that, from the perspective of a language-endowed social intelligent agent, processing metaphor – and figurative language in general – is an epiphenomenon of an agent's ability to learn lexical senses on the fly and to postpone or disregard analysis of parts of input without compromising its functioning. This work is a step toward developing an explanatory theory of why people often operate successfully without fully understanding language input and sometimes without even attempting to process parts of it. The basis for understanding novel metaphors is modeling the life-long, "diachronic" functioning of an agent. This paper provides a descriptive account of this theory and discusses how several existing components of the OntoAgent environment are used or are being extended to support implementation of the theory in a proof-of-concept system.

1. Introduction

Metaphor is a fascinating and amply studied phenomenon. It has been addressed from a broad variety of premises and in different contexts: in rhetoric since Aristotle, in literary criticism (e.g., Skulsky, 1986), semiotics (e.g., Eco, 1979), a variety of schools in linguistics (e.g., Lakoff, 1993; Steen, 2007), psychology (e.g., Bowdle and Gentner, 2005), psycholinguistics (e.g., Glucksberg, 2003), philosophy (e.g., Bayler-Jones, 2009; Lepore and Stone, 2010) and neuroscience (e.g., Goldstein et al., 2012).

For about the past five years we have been witnessing a new wave of research on metaphor in the computational linguistics (CL) community. This work follows the standard methodology adopted in practically all CL work of the past 20 years which involves:

- choosing a clearly demarcated language phenomenon such as modality, multi-word expressions, reference resolution or word sense disambiguation
- selecting a subset of its manifestations, typically constrained to examples that are most readily handled by knowledge-lean processing methods
- formulating a "task" or "task definition" to treat this subset
- preparing textual resources, particularly annotated corpora, to be shared by the community
- developing algorithms for solving the task using a combination of standard statistical processing methods, the specially developed resources, and other available resources
- evaluating the algorithms: preferably on shared unseen texts, preferably as part of a competition among many research groups, and often on texts that have been manually pre-annotated.

Though the above methodology has not yet been fully implemented in the new CL metaphor community, doing so is one of its stated goals. Thus Shutova (2015) writes: "So far, the lack of a common task definition and a shared data set have hampered our progress as a community… This calls for a unification of the task definition and a large-scale annotation effort that would provide a data set for metaphor system evaluation…" (p. 617).

Adopting this methodology has a lot to recommend itself: it concentrates efforts on processing a selection of important phenomena, fosters research community building, and facilitates a fruitful mixture of collaboration and competition among research teams. It is equally important that it creates an atmosphere of excitement and fun and instills a sense of purpose and satisfaction. Indeed, the main results of such work are numbers reflecting the percentage of decisions that systems made correctly. Over time, algorithms are expected to improve and the percentages are expected to grow. This offers a tangible measure of progress and kindles hopes of reaching a 100% solution on the content of the task definition.

Time will tell whether the above methodology will actually fulfill the promise of a 100% solution for a selective task or whether it is extensible to tasks defined in a more comprehensive way to account for representative inventories of language phenomena. A detailed analysis of the advantages and shortcomings of this methodology is outside the scope of this paper. Instead, we will name just a few considerations relevant to our argument: a) under the above methodology, the selected phenomena are seldom, if ever, studied in all their complexity; b) known constraints on the quality of annotation strongly impact (constrain) the selection of phenomena to be included in the task definition; c) evaluation measures are often not truly informative, both because of the selectivity of phenomena treated and the tendency to evaluate systems on manually annotated datasets; d) by concentrating on narrowly defined tasks, the methodology fosters a disconnect from the greater aim of developing integrated throughput-oriented application systems; and, last but not least, e) from the point of cognitive science – or even science in general – a major shortcoming of this approach is its intrinsic lack of explanatory power; developing explanatory theories is not among the goals of modern computational linguistics.

In this paper we put forward a theoretical and methodological alternative to the above trend. Taking into account a variety of issues discussed in the approaches to metaphor accumulated in philosophy, psychology and linguistics, as well as the system-building experience of early AI, we present an initial sketch of an explanatory theory of metaphor interpretation. This theory:

- argues that metaphor interpretation does not deserve its unique status in language understanding; instead, it **treats metaphor within a general approach to handling "unexpected" input** that arises due to various ostensible anomalies in language use
- concentrates on **automatic detection and interpretation of metaphors** (and other "anomalies") not by people but **by cognitive systems**, specifically, by language-endowed social artificial intelligent agents (LEIAs)
- focuses on cooperative, task-oriented collaborations in which human-agent teams are expected to dynamically model their collaborators' mental states using a **theory of mind**
- uses **a large inventory of knowledge sources** that provide heuristics for the reasoning and decision-making necessary for processing anomalies, including metaphors
- argues that both people and social artificial intelligent agents can very often **successfully function with incomplete understanding** of some residual anomalies in communication
- views the interpretation of any input not as a one-off operation but, rather, as a part of the lifelong operation, including **lifelong learning,** of LEIAs.

We will now discuss the classification of anomalies, including but not limited to metaphors, and their treatment by LEIAs in OntoAgent.

2. Classification of Anomalies

Our theory of anomaly treatment is an extension of the theory of Ontological Semantics (Nirenburg and Raskin, 2004) into the realm of computational cognitive modeling. This work is under continuous development in the OntoAgent cognitive agency framework. To understand what agents consider an anomaly, one must first understand the basic process of language understanding in OntoAgent.

An OntoAgent uses its knowledge about language, stored in its lexicon, and its knowledge about the world, stored mostly in its ontology, to automatically generate ontologically-grounded text meaning representations (TMRs). TMRs are stored in an agent's memory and serve as input to reasoning and decision-making (McShane and Nirenburg, 2012). The core operations in an OntoAgent's semantic analyzer are lexical disambiguation and semantic dependency determination, which are carried out together and rely on the knowledge recorded in the lexicon and the ontology (McShane et al., forthcoming). The lexicon records the linked syntactic and semantic expectations of argument-taking words, including multi-word expressions, whereas the ontology lists the semantic constraints on the properties of the objects and events that realize the meanings of lexical items. For example, LEIA's lexicon indicates that the main sense of *eat* is optionally transitive and means INGEST. In the basic diathesis, the meaning of the subject fills the AGENT case-role of INGEST and the meaning of the direct object fills its THEME case-role. The ontology indicates that the AGENT of INGEST must be an ANIMAL and the THEME should be an INGESTIBLE (i.e., food, drink or an ingestible medication). In the simplest, optimal case, an input meets these expectations – as in *Frank ate a slice of pizza* – and the analyzer generates exactly one interpretation. However, there are two other possible outcomes, which we call the *residual ambiguity anomaly* and the *unexpected input anomaly*, as shown in Figure 1.[1]

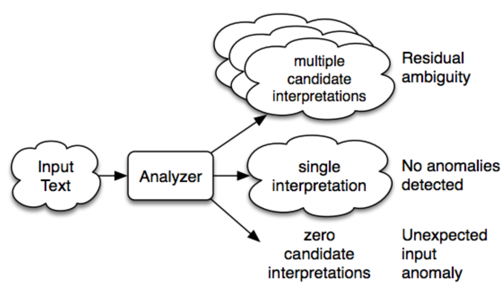

Figure 1. Outcomes of semantic analysis.

A **residual ambiguity anomaly** occurs if semantic constraints alone are not sufficient to choose between several candidate readings: e.g., in the sentence *Pirates should be sent to jail*, the meaning of *pirates*—be they pirates at sea or illegal copiers of software, etc.—cannot be disambiguated without further context, since they are both HUMANs and any HUMAN can be the THEME of IMPRISON.

An **unexpected input anomaly** arises in one of two cases: a) when the constraints on known word senses are *too tight*, so that no legal combination of word senses can be found; or b) when the lexicon is *incomplete*, lacking words or word senses providing constraints that would match. For example, the ontological expectations about INGESTing will be violated in inputs like *The toddler ate a candle* or *The conjurer ate a*

[1] In this paper we do not address pre-semantic anomalies, such as non-canonical syntax, or residual ambiguities that require reasoning about the larger context, as in our pirate example.

plate of glass shards. Initial methods for dealing with all of these types of anomalies were reported in Nirenburg and Raskin (2004). This paper extends those approaches and adapts them to the needs of a social LEIA participating in dialogs.

The knowledge needed to generate a single, high-confidence semantic analysis can be missing for many reasons. LEIA lexicons will never be complete – a good LEIA will have at least to try to make sense of something like Lewis Carroll's *Jabberwocky*. Alternatively, the system's morphological analyzer may fail to detect the overtly recorded lemma for an inflectional form of an input word (*corpora* > *corpus*). A LEIA's lexicon may also be missing not a word but a word sense. For example, it may lack the second sense of *guzzle,* as listed in The American Heritage Dictionary: "To consume to excess: *a car that guzzles gas.*" Without this sense, the LEIA will encounter a violation of selectional restrictions in the example *My car guzzles gasoline*. In early AI approaches (e.g., Carbonell, 1982; Fass and Wilks, 1983), such selectional constraint violations served as the trigger for detecting and resolving metaphor, a tradition kept alive both in later knowledge-based approaches (e.g., Wilks et al., 1996) and in the new distributional-semantic CL paradigm. But the traditional treatment strategy is unnecessarily narrow on two fronts. First, the apparent metaphor might best be added as a new word sense to the lexicon, rather than being perpetually treated as a metaphor (see discussion below). Second, unexpected input anomalies can signal tropes other than metaphors, such as the metonymy in *Richter played Brahms*.[2]

Although to this point we have concentrated on violating selectional restrictions on verbs, text inputs can violate other types of ontological constraints as well. For example, *Lawyers are sharks* will contradict subsumption relations in the ontology: the corresponding concepts are on different branches ascending to the ontological concept ANIMAL. Similarly, if we consider that the main meaning of *consists of* is HAS-AS-PART, then *effort* violates expectations in the input *A good jam consists of fruit, sugar, water and effort*, since *effort* does not refer to a physical object. Finally, the meaning of a modifier can fail to unify with any available meaning of the modified, as in *the brown company*, an indirect reference to United Parcel Service.

Sometimes detecting an anomaly requires world knowledge that extends far beyond the speech context. For example, a LEIA will miss the following joke if its lexicon does not include a sense of *comfortable* meaning, roughly, *not impecunious*: *Crossing 42nd Street, Shapiro is hit by a car. While he's lying flat on the pavement, people rush toward him and somebody asks: "Are you comfortable?" "I make a living*," *says Shapiro*. It will also fail to recognize the humor in the following if it doesn't know that a sore throat can cause one to lose one's voice, and that extra-martial affairs are typically kept secret: *A patient with a sore throat knocks on the door of a rural medical practice. The physician's wife answers the door. "Is the doctor home?" rasps the patient in a coarse whisper. "No, he isn't. Come in!"*

In addition to a lack of knowledge, a lack of reasoning ability may also inhibit the detection of anomalies. Given the input, *I told you a million times, don't slouch!*, a LEIA will be able to detect an anomaly – hyperbole – only if it understands that a million iterations is unlikely. Jumping ahead a bit, this example illustrates a case where a LEIA may decide to forgo a complete analysis of input by recognizing that the meaning of *I told you a million times* is not central to the meaning of the command *Don't slouch!*

Classifications of anomalies that are created for human consumption mostly define familiar tropes such as metaphor (*His gaze shot through me*), metonymy (*play Brahms*), zeugma *("You*

[2] One way of resolving this would be to interpret *Brahms* as an elliptical metonym for *a composition by Brahms* (e.g., Fass, 1997).

are free to execute your laws, and your citizens, as you see fit," Star Trek: The Next Generation), synecdoche (*All hands on deck!*), hyperbole (*I've told you a million times…*), litote (*Michelangelo was an above-average sculptor*) and irony (*I just love cleaning the house*). Such classifications might be useful for computational applications a) if it were possible to detect automatically which inputs belonged to which categories, and b) if detecting the category impacted the choice of interpretation algorithm. However, we have concluded that separating the processing of metaphor from other anomalies is not realistic for LEIAs. Moreover, it does not seem to be *a priori* necessary to identify an anomaly as a metaphor before processing it, though such identification may be desirable *after* anomaly processing to contribute to the LEIA's updating of its theory of its interlocutor's mind. (E.g., when talking to a person who uses extensive figurative language, a LEIA might choose to generate figurative language as well, a capability we will explore separately.) Our conclusion is that interpretation of metaphor and figurative language in general is an epiphenomenon of processing unexpected input anomalies.

3. An Inventory of Methods for Resolving Semantic Dependency Anomalies

Since the knowledge resources of a LEIA can never be complete, and since language use is endlessly open-ended, LEIAs expect to regularly encounter both the residual and unexpected input anomalies shown in Figure 1. They also expect to fail to recognize some anomalies, e.g., interpreting *Old Mr. Jones kicked the bucket yesterday* in its non-idiomatic meaning if the idiomatic meaning is recorded in the lexicon.

Developing methods for resolving residual ambiguity and unexpected input anomalies has been one of the major topics of research in Ontological Semantics for many years. A number of methods have been, and continue to be, experimented with in various implementations of the theory (see McShane et al. 2016 for a comparison among three implementations). We classify each method as *static* or *dynamic* and *synchronic* or *diachronic*, as shown in Table 1.

Table 1. OntoAgent methods for resolving anomalies. *Static* methods use and extend the system's knowledge resources. Dynamic methods extend system's algorithms to handle anomalies. Synchronic methods process sentences in isolation. Diachronic methods operate at the discourse level and extend to the lifetime of an agent. References for implemented methods and development status are shown in *italics*.

	Synchronic	**Diachronic**
Static	1. Multivalued selectional constraints *Onyshkevych and Nirenburg 1991*	6. Learning by reading *Nirenburg et al. 2007*
	2. Lateral selectional constraints *Mahesh et al. 1997*	7. Learning through dialog (by being told) *Nirenburg et al. 2010 & under development*
Dynamic	3. Dynamically tightening / relaxing selectional constraints *Mahesh et al. 1996*	8. Using the discourse situation to resolve anomalies *Under development*
	4. Unilateral use of selectional constraints *McShane and Nirenburg 2002*	9. Delaying resolution of anomalies *McShane and Nirenburg 2015 & under development*
	5. Using the ontology as search space *Onyshkevych 1997, 1998*	10. Actionability halts analysis *McShane and Nirenburg 2015 & under development*

Static methods concentrate on extending the system's knowledge bases, whereas *dynamic* methods extend the basic processing algorithms to better handle anomalies. Using *synchronic* methods, decisions are made based on knowledge that is immediately available, whereas using *diachronic methods*, the resolution of anomalies can be extended over time. Space constraints allow for but the briefest survey of the above methods. We start with the synchronic ones, which were originally developed for a non-agent implementation of Ontological Semantics that operated at the sentence level, but are equally useful to LEIAs. We then turn to diachronic methods, which are specific to LEIAs operating over larger discourse segments and time periods.

1. ***Multivalued selectional constraints*** (Onyshkevych and Nirenburg, 1991) [Synchronic, Static]. The OntoAgent ontology offers three levels of semantic constraints for property values: *default* if there is a specific, most common filler; *sem* for the basic filler(s); and *relaxable-to* for rare but possible fillers. For example, the AGENT for SURGERY is listed as *default: SURGEON, sem: PHYSICIAN, relaxable-to HUMAN.* The basic semantic analysis algorithm matches the semantics of each of the lexical senses of a predicate with that of each of the senses of its arguments. The matches are rated, with a match on the DEFAULT facet adding a bonus to the rating and a match on RELAXABLE-TO adding a penalty relative to a match on SEM. The reading with the highest cumulative rating at the proposition level is chosen for the nascent TMR. The DEFAULT facet supports the elimination of residual ambiguity, in expectations that tightened constraints will rule out spurious candidates. For example, resolving the reference of *he* in the following sentence requires knowing that a surgeon is the typical agent of surgery: *The boy's father talked with the surgeon before he operated on him.* The RELAXABLE-TO facet was introduced to support the resolution of anomalies. Let us return to the example of eating, analyzed using the ontological concept INGEST. Its THEME on the *sem* facet is INGESTIBLE, which covers food, drinks and ingestible drugs. However, our example was *The toddler ate the candle*. We know that children, not to mention animals, can swallow all sorts of things, so ideally the ontology will indicate that the THEME of INGEST on the *relaxable-to* facet is the set of materials or physical objects that are of a swallowable size, shape, temperature and consistency. This level of specification would support a perfect analysis by the system. However, formally specifying this set of unusual edibles requires a lot of work, particularly as judged against the rarity of such inputs; so, highly specified constraints for *relaxable-to* are often missing. A stopgap in acquisition is to list a coarse-grained filler for *relaxable-to*, such as PHYSICAL-OBJECT. Although this would permit the system to correctly analyze our toddler-with-candle example, it would cause the agent to miss the intended anomaly of an input like *The dog ate my car!,* which certainly does not mean that he ingested the whole car – he probably just scratched it up a bit with his teeth.

2. ***Lateral selectional constraints*** (Mahesh et al., 1997) [Synchronic, Static]. Lateral selectional constraints exist between the concepts serving as case-roles of events. For example, although the concept INGEST is described as having an ANIMAL as its AGENT and an INGESTIBLE as its THEME, we know that birds eat worms, wolves eat meat, and horses eat oats. These pairs of case-role fillers can be recorded in the ontology and used for subsequent agent reasoning. So if a LEIA encounters the input *The horse [unknown-word] some oats*, it can guess that the unknown word means INGEST or something closely related to it through an ontological script, such as CHEW, SWALLOW or DIGEST.

3. ***Dynamically tightening or relaxing selectional constraints*** (Mahesh et al., 1996) [Synchronic, Dynamic]. A LEIA can dynamically tighten or relax recorded ontological constraints on the basis of contextual information. Consider the example *This restaurant makes excellent lasagna.* Alt-

hough the input uses the quite generic verb *make*, we as humans understand that lasagna is actually baked. If the agent wants to successfully disambiguate the light verb *make,* as well as better understand what type of activity was undertaken, it can search the ontology and learn the following: a) BAKE is a filler of the THEME-OF property of the ontological concept LASAGNA (along with, e.g., INGEST); b) BAKE is an ontological descendent of PREPARE-FOOD that is the meaning of one sense of *make*; and c) FOOD-ORGANIZATION, the meaning of *restaurant*, matches the selectional constraint on the AGENT of BAKE (on the RELAXABLE-TO facet). So, interpreting *make* as BAKE works perfectly.

4. ***Unilateral use of selectional constraints*** (McShane and Nirenburg, 2002) [Synchronic, Dynamic]. Consider another instance of unexpected input: *An [unknown-word] was eating loudly.* Even though only one case-role filler is available, the agent can hypothesize that *unknown-word* refers to some sort of ANIMAL since the subject of the verb *eat* in the active voice should be the AGENT of INGEST, and the AGENT of INGEST is always an ANIMAL.

5. ***Using the ontology as a search space*** (Onyshkevych, 1997, 1998) [Synchronic, Dynamic]. Consider the input *The big glasses borrowed my bike,* which includes a metonymy for *the person wearing big glasses.* The LEIA will recognize that neither lexically available interpretation of *glasses* – either SPECTACLES or a set of objects of the type DRINKING-GLASS – is a valid filler of the AGENT slot of BORROW. So it will use the ontology as a search space to try to figure out how SPECTACLES or a set of DRINKING-GLASSes could represent the needed agent. It will compute the weighted distance between the expected agent, HUMAN, and both of these concepts. The cumulative score for each reading will be a function of the length of the path and of the cost of traversing each particular relation link. Both the traversal costs and the combination of evidence function were induced empirically. The system was trained using simulated annealing on the 40 most frequently used ontological properties.

6. ***Learning by reading*** (Nirenburg et al., 2007) [Diachronic, Static]. Learning by reading is triggered after synchronic methods have done what they can to interpret an unknown lexical item. The LEIA searches a corpus for sentences containing this item, uses its analyzer to generate TMRs for them using the abovementioned methods for processing unknown words, then uses the information about the newly acquired word available in the TMRs to appropriately constrain the fillers of its ontological properties. For example, given the examples *The red yool$_{UNKNOWN}$ ate a carrot*, and *A black yool$_{UNKNOWN}$ ran up the tree,* the LEIA can learn that *unknown-word* is some type of ANIMAL (from the unilateral selectional constraints on the AGENT of INGEST and RUN) and that this type of ANIMAL can have at least "red" and "black" as its values for COLOR.

7. ***Learning by being told*** (Nirenburg et al., 2010) [Diachronic, Dynamic]. We draw examples of learning by being told from the Maryland Virtual Patient system, in which dialog-enabled virtual patients are diagnosed and treated by people playing the role of clinicians in training. If the doctor says, "You have *unknown-X*," the virtual patient can use a combination of lexical knowledge and contextually-triggered expectations to guess that X is some sort of disease; accordingly, it learns the word *unknown-X* and maps it to the concept UNKNOWN-X, which is generated as a child of DISEASE. Similar reasoning can be used to learn about medical procedures from inputs like "I think you should have a *unknown-Y"*, and to learn the properties of ontological concepts from inputs like "*Unknown-Y isn't dangerous and doesn't hurt much.*"

8. ***Using the discourse situation to resolve anomalies*** [Diachronic, Dynamic]. We are currently working on using as a source for disambiguation heuristics the TMRs for prior inputs in a dialog or text stored in the agent's short-term memory. This is our way of operationalizing the familiar premise that context helps disambiguation, and it relies on the co-occurrence of word senses. For example, the input *Our tree sings to us* contains a sortal incongruity, meaning that either *tree* or *sing* is being used non-literally. Reference resolution can help to decide which one. For example, given the context, *We planted a beautiful oak tree in the backyard and now our tree sings to us every night,* the LEIA will analyze the first instance of *tree* as the physical object TREE, which can, indeed, be planted, and it will establish a coreference relationship between the two instances of *tree*. This grounding of the meaning TREE in the second clause suggests that *sing* is the lexeme being used non-literally. Such reference-based grounding can also be leveraged *after* the anomalous clause, as would be the case for the input *Our tree sings to us every night. We planted it in the backyard just last year.*

9 & 10. ***Delaying resolution of anomalies & using actionability judgments to halt analysis*** (McShane and Nirenburg, 2015) [Diachronic, Dynamic]. Most approaches to automatic language processing, irrespective of the methods they use, have as their ultimate goal a complete and correct analysis of language inputs. We hypothesize that people behave differently: they analyze only as much of the input, and only as deeply, as they deem necessary to react in reasonable ways. Accordingly, the core distinguishing feature of our theory of textual anomaly processing is enabling the agent to decide to disregard or incompletely resolve some ambiguities and anomalies. While this will allow the agent to get on with more important things in many cases, the agent also runs the risk of misunderstanding and concomitant suboptimal action; so it must also have the ability to return to previously unresolved or "underresolved" ambiguities and anomalies if it detects that a prior decision was incorrect. Our approach to delayed and incomplete resolution of language meaning is guided by the following four hypotheses.

Hypothesis 1. *People do not always pursue the most highly specified interpretation of an input.* For example, consider the elasticity of the meaning of *good*. *This car is good* may mean that it is comfortable, reliable, fast, economical, or any combination of the above. In the absence of a special reason for extra precision, on receiving this input, most people will be satisfied by understanding that the speaker likes the car in question. This hypothesis dovetails with the views of Lepore and Stone (2010), who argue that metaphorical meaning does not need to be fully semantically interpreted or recorded.

Hypothesis 2. *People sometimes stop processing an interlocutor's dialog turn before its end.* This can happen during the incremental processing of an input for at least two reasons. First, the hearer can decide that he or she has a good enough idea of what is to come to interrupt with a response, as in the following exchange from the Switchboard corpus (Godfrey et al. 1992): *A.: That would have meant a total attack of Iraq within, you know, three hours of when the weapon was shot. B.: Oh, so you think it was fear that kept Iraq from – A.:* ***Right****. B.: – using it.* The second reason people fail to wait for the conclusion of the interlocutor's statement is that what the speaker has said so far is already *actionable* – i.e., it fulfills all of the prerequisites for triggering a plan to attain an active goal (McShane and Nirenburg 2015). For example, in the Maryland Virtual Patient application, the agent is ready to generate a response after processing just the first sentence in the following input from the doctor: *Do you have chest pain? That's the $64,000 question at this point in time*. Depending on how turn-taking is implemented, the agent could either interrupt the doctor after the first sentence, or wait for – but essentially ignore the meaning of – the 2nd sentence before answering the question.

Hypothesis 3. *People often do not attempt to resolve an ambiguity or anomaly immediately* after they encounter it, since subsequent input might well offer a clarification. For example, *President Taft was an elephant* is anomalous, with elephant metaphorically referring to weight (heavy), memory (excellent) or political party membership (Republican). The three readings would be disambiguated by the respective rejoinders, *He weighed over 350 pounds, He never forgot anything he read*, or *Everybody knows he was a Republican* (this example is inspired by Hobbs, 1992). This also seems to be a tactic people use while speaking a foreign language they know only moderately well.

Hypothesis 4. *People selectively record results of anomaly resolution as tentative new lexical senses*, improving the precision of, and confidence in, these new senses as a side effect of subsequent functioning. This implies that some results of anomaly resolution are not deemed worthy of remembering. This hypothesis reflects the notion of "the career of the metaphor" (Bowdle and Gentner, 2005), though applied to an individual agent rather than to an entire language community.

4. A Sequence of Algorithms for Resolving Semantic Dependency Anomalies

Figure 2 shows the manner in which we integrate the methods of Table 2 into an algorithm for resolving semantic dependency anomalies. It focuses squarely on anomaly treatment, not detailing, for reasons of space, many relevant aspects of agent functioning, such as dynamically maintaining a belief repository, managing an agenda of goals and plans, or modeling the mental state of collaborators/interlocutors – all of which we discuss elsewhere (e.g., Nirenburg and McShane, 2012, McShane et al., 2013). The algorithm is at the moment constrained to just a subset of control options: it would be more realistic if an agent could at any moment choose to stop pursuing further, deeper processing of the input or, alternatively, ask its collaborator for clarification (we do intend to introduce these options in the future).

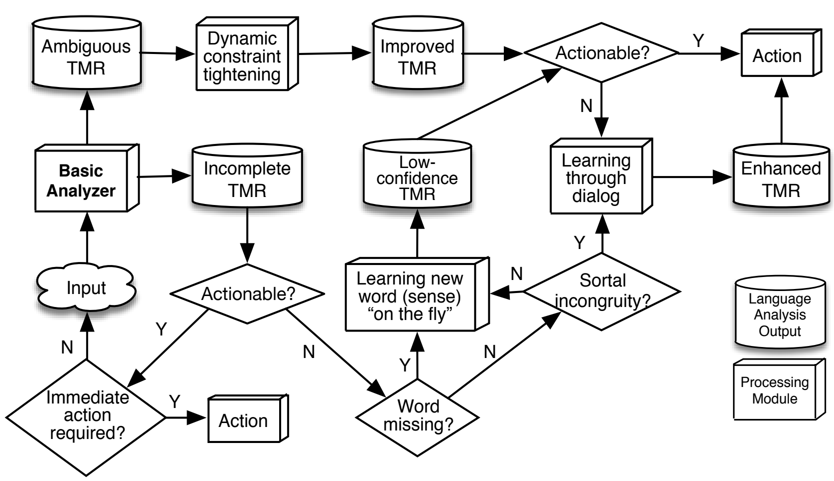

Figure 2. Processing residual ambiguities and unexpected input in OntoAgent.

Let us walk through Figure 2. The two types of anomalies in question are residual ambiguity, resulting in more than one candidate TMR, and unexpected input, resulting in an incomplete TMR.

In the case of residual ambiguity, the LEIA will undertake dynamic constraint tightening which, in the best case, will result in an improved TMR: e.g., *make lasagna* will be understood as BAKE (THEME LAZAGNA). If this interpretation is actionable, the agent takes action. Otherwise, it asks its human collaborator for clarification through dialog, as might be needed for the input *She finished the wall,* in which case the nature of the event – painting, building, decorating? – cannot be determined by ontological defaults. Human clarification will result in a TMR that is actionable.

In the case of unexpected input, which results in an incomplete TMR, the agent will decide if its partial understanding is actionable. If it is, and if immediate action is required, it takes action: *First douse the fire and then [garbled input]!* If action is possible but not urgent, then the agent postpones action, waiting for further input which might provide the necessary information to resolve what was previously unclear: *I recommend having a Heller myotomy.* [I don't know what that is. Let me wait and see.] *It is a surgical procedure to the esophagus.* If, by contrast, an incomplete TMR is not actionable, that might be due to an unknown word or sortal incongruity (e.g., *The toddler ate the candle*). In the case of an unknown word, the LEIA will attempt learning by reading, whose result might be actionable or might require further elucidation by the human interlocutor. In the case of sortal incongruity – which can, in some cases, be due to a metaphorical or other trope-based usage of a known word – the agent must ask the interlocutor for help. This "learning by being told" can result in a new word sense being added to the lexicon, a new filler being added to the given ontological property (typically on the *relaxable-to* facet), or no amendments to the static knowledge sources if the usage is idiosyncratic and not expected to be encountered again. Remember, recording the results of anomaly processing in the lexicon and ontology is not without future cost, since it may hinder ambiguity resolution in the future. That is, if conventional metaphors are treated as word senses, there will be many more word senses in the system's lexicon from which to choose. A judgment must be made in each case whether the selectional constraints and other properties of the meaning of the new word sense distinguish it from other senses sufficiently for the analyzer not to end up with having to eliminate residual ambiguity anomalies in the future.

Let us summarize, using the terminology of Table 2, the five kinds of TMRs that can result from basic text analysis:

- Optimal, *non-anomalous* TMRs;
- *enhanced* TMRs, which are high quality thanks to explanation by the human interlocutor;
- *improved* TMRs, which started out being ambiguous but, as a result of agent reasoning, became either less ambiguous or unambiguous;
- *incomplete* TMRs, for inputs with either an unknown word or a sortal incongruity; and
- *low-confidence* TMRs resulting from the agent's attempt to learn by reading, which is a complex, error-prone undertaking.

The above can be functionally classified into two groups based on how they are integrated into agent memory. The highest quality TMRs – non-anomalous ones and those enhanced thanks to human input – are directly incorporated into the agent's long-term memory. By contrast, when the agent produces lower-quality TMRs – improved, incomplete and low-confidence ones – it will attempt to improve both the TMRs and the relevant aspects of its knowledge bases before committing the TMRs to long-term memory. The most independent action is learning by reading; but, if that is unsuccessful or incomplete, learning through dialog can also be undertaken. Of course, both of these will be undertaken only if the agent determines that the input is sufficiently important to merit the processing time and/or human effort. Modules for both of the types of learning have been implemented and demonstrated, on a limited scale, in OntoAgent, as was a module for augmenting the agent's belief repository. Extending and improving these learning algorithms is a central direction of our team's current and future work.

4. Closing Thoughts

The distinction between conventional and novel metaphors has been firmly established in linguistics (e.g., Nunberg, 1987) and psychology (e.g., Gibbs, 1984). Bowdle and Gentner (2005) view the novel-conventional metaphor continuum in an etymological perspective and argue that metaphors conventionalize and diachronically lose their metaphoricity. Most metaphors discussed within the popular conceptual metaphor theory (e.g., Lakoff, 1993) are actually conventional and, therefore, presumably exist in a native speaker's lexicon. Even if the early AI approaches to metaphor do not state it overtly, their underlying motivation was to use metaphor processing as a means of bypassing the need for lexical and conceptual knowledge acquisition. In a recent survey of work on metaphor in computational linguistics Shutova states: "Much of the metaphor processing work has focused on conventional metaphor, though in principle capable of identifying novel metaphor as well" (2015, p. 582).

Theorists go beyond the novel/conventional distinction. Steen (2011) introduces a distinction between deliberate and non-deliberate metaphors. But as he concedes, "the processes leading up to the product of metaphor comprehension […] are largely immaterial to the question of whether their product counts as a deliberate metaphor or not" (ibid: p.85). This corroborates our position: to successfully process input containing conventional metaphors the hearer does not need to realize that a metaphor is present. Conventional metaphor *qua* metaphor may be of interest to scholars or as the subject of an entertaining etymological parlor game. But to understand *ballpark* in *ballpark figure* it is not necessary to know that it is a (baseball) metaphor. It is appropriate then to pose the question of why AI and CL metaphor researchers insist on including conventional metaphor in the purview of their systems.

We hypothesize that people usually process novel and deliberate (Steen, 2011) metaphors in the same manner in which they process unknown lexical units that are not metaphorical – by learning their meaning over time from their use in text and dialog, and recording those meanings in their lexicons for later use. In other words, the novel (non-metaphorical) senses of *pocket* and *bank* in *He pocketed the ball by banking it off two rails* will be learned with the help of knowledge of the domain (billiards) and general knowledge of what can typically be done with a billiard ball. By the same token, the meaning of *albatross* in, say, *O'Malley's heaviest albatross is the state of his state* will also be understood based on the hearer's knowledge of the overall context, with no need for the hearer to have read, or even know about the existence of, Coleridge's *The Rime of the Ancient Mariner*.

Of course, building an agent that models an etymologist is a potentially interesting research direction, but it is much more important in agent systems to cover conventional metaphors. And we argue that the best way of doing this is by viewing the task as a routine part of the lifelong enhancement of an agent's knowledge resources. An agent of this kind will fail to register the esthetic contribution of an extended metaphor like the following, but this is equally true about many people – after all, not everybody knows about baseball: [A team leader cajoling a team member] Eric, the *bases are loaded*; tomorrow's demo is crucial. Please stop *grandstanding* and *playing hardball*, *step up to the plate*, join the effort and *lead off* with a *ballpark figure*.

Initially, some readers might not have fully understood the meaning of the title of this paper.

However, most everyone will have guessed that the authors intended to say something negative about the study of metaphor.[3] Some readers will also have understood that the authors would justify this attitude on the grounds that the study of metaphor is unnecessary from some point of view. Having read on, readers who still remember the title would realize what it intended to convey – that separating metaphor detection and interpretation from the treatment of other types of figurative language and other semantic anomalies violates the dictum "entities must not be multiplied beyond necessity." Now, readers (such as LEIAs) with no training in philosophy may have recognized *Occam* as a named entity without realizing that *Occam's razor* refers to the above dictum. Such readers would fully understand this paper's title only after having read the previous sentence. The above observations further motivate our contention that delayed interpretation of input is a viable and potentially effort-saving strategy for agents.

Some readers will also appreciate the *double entendre* in the title due to the metaphorical use of an action (*slashing*) associated with a physical tool (*razor*) that once served as the source of the metaphor to describe the mental tool (*Occam's razor*) of the title. While recognizing this may be a nice bonus, it is not essential for understanding the main argument of the paper. This observation illustrates and motivates our contention that agents can often function optimally without understanding all of an input. Anybody who has ever communicated in a foreign language can vouch for this – it is common practice to get by in speech situations without full understanding of all lexical material. Sometimes this leads to misunderstandings or embarrassment but, more often than not, it works well enough to achieve success in communication. Of course, the $64,000 question is how to teach LEIAs to determine what, if any, parts of an input they can disregard with impunity. This is one of the directions of our team's future work.

Finally, our agents model the human ability to constantly learn new lexical material and new facts. With respect to processing figurative language, this allows us to recreate in the system's ontogeny the phylogeny of a linguistic community – our agents will be conventionalizing metaphors and other tropes in the regular course of their operation.

Acknowledgements

This research was supported in part by Grants N00014-09-1-1029 and N00014-16-1-2118 from the U.S. Office of Naval 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.

References

Bayler-Jones, D. (2009). *Scientific Models in Philosophy of Science*. University of Pittsburgh Press.

Bowdle, B. & Gentner, D. (2005). The career of metaphor. *Psychological Review*, 112: 193–216.

Carbonell, J. (1982). Metaphor: An inescapable phenomenon in natural language comprehension. In W. Lehnert and M. Ringle (eds.), *Strategies for Natural Language Processing*. Erlbaum, pp. 415–434.

Eco, U. (1979). The semantics of metaphor. In *The Role of the Reader: Explorations in the Semiotics of Texts*. Bloomington: Indiana University Press, pp. 67–89.

[3] We assume that no reader will fail to interpret *metaphor* in the title as a metonym for *the study of metaphor*.

Fass, D. (1997). *Processing Metonymy and Metaphor*. Greenwich, CT: Ablex Pub.

Fass, D. & Wilks, Y. (1983). Preference semantics, ill-formedness, and metaphor. *American Journal of Computational Linguistics*, 9(3): 178–187.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences* 7(2): 92–96.

Godfrey, J., Holliman, E., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing,* Volume 1, pp. 517–520, Washington, DC.

Goldstein, A., Arzouan, Y. & Faust, M. (2012). Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization. *Brain and Language* 123: 137–142.

Hobbs, J. R. (1992). Metaphor and abduction. In A. Ortony, J. Slack, and O. Stock, editors, *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*, pp. 35–58. Springer.

Lakoff, G. (1993). The contemporary theory of metaphor. In Ortony, A. (ed.), *Metaphor and Thought* (2nd ed.). New York: Cambridge University Press, pp. 202–251.

Lappin, S. (1981). *Sorts, Ontology and Metaphor*. De Gruyter.

Lepore, E. & Stone, M. (2010). Against metaphorical meaning. *Topoi* 29(2): 165–180.

Mahesh, K., Nirenburg, S. & Beale, S. 1996. Exploiting dynamic contexts for word sense disambiguation. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*.

Mahesh, K., Nirenburg, S. & Beale, S. (1997). If you have it, flaunt it: Using full ontological knowledge for word sense disambiguation. *Proceedings of Theoretical and Methodological Issues in Machine Translation (TMI-97)*, Santa Fe, NM.

McShane, M. & Nirenburg, S. (2002). Reference and ellipsis in ontological semantics. Memoranda in Cognitive and Computer Sciences MCCS-02-329, CRL, New Mexico State University.

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

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., & Beale, S. (Forthcoming). Language understanding with Ontological Semantics. *Advances in Cognitive Systems*.

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

Nirenburg, S. & McShane, M. (2012). Agents modeling agents: Incorporating ethics-related reasoning. *Proceedings of the Symposium on Moral Cognition and Theory of Mind*, Birmingham, UK.

Nirenburg, S., McShane, M., Beale, S., English, J. & Catizone, R. (2010). Four kinds of learning in one agent-oriented environment. *Proceedings of the First International Biologically-Inspired Cognitive Architectures Conference*.

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.

Nunberg, G. (1987). Poetic and prosaic metaphors. *Proceedings of the 1987 Workshop on Theoretical Issues in Natural Language Processing*, pp. 198–201.

Onyshkevych, B. (1997). An Ontological-Semantic Framework for Text Analysis. PhD Thesis, Carnegie Mellon University. Technical Report CMU-LTI-97-148.

Onyshkevych, B. (1998). Nominal metonymy processing. *Proceedings of the ACL-98 Workshop on the Computational Treatment of Nominals*.

Onyshkevych, B. & Nirenburg, S. (1991). Lexicon, ontology and text meaning. *Proceedings of the ACL Workshop on Lexical Semantics*. Berkeley, CA, June.

Shutova. E. (2015). Design and evaluation of metaphor processing systems. *Computational Linguistics*, 40: 579–623.

Skulsky, H. (1986). Metaphorese. *Noûs*, 20(3): 351–69.

Steen, G. (2011). From three dimensions to five steps: The value of deliberate metaphor. *Metaphorik.de*, 21: 83–110.

Wilks, Y., Barnden, J. & Wang, J. (1996). Your metaphor or mine: Belief ascription and metaphor interpretation, pp. 945–950. In Partee, B. H. & Sgall, P. (eds.). *Discourse and Meaning*. John Benjamins.