Advances in Cognitive Systems 2 (2012) 95-112 Submitted 09/2012; published 12/2012

Resolving Elided Scopes of Modality in OntoAgent

Marjorie McShane MARGE@UMBC.EDU
Sergei Nirenburg SERGEI@UMBC.EDU
Stephen Beale SBEALE@UMBC.EDU
Ben Johnson BENJ1@UMBC.EDU
Computer Science and Electrical Engineering, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250 USA

Abstract

This paper describes how intelligent agents modeled within the OntoAgent cognitive architecture treat the linguistic phenomenon of modal scope ellipsis. The approach offers two innovative aspects with respect to cognitive modeling: (1) ellipsis treatment is distributed across processing modules in a psychologically plausible way and (2) agents are prepared to incorporate calculations of utility into their ellipsis resolution efforts. The latter means that agents can evaluate their confidence in each ellipsis resolution decision and, in cases of low confidence, determine whether or not it is worthwhile to pursue a clarification. Endowing agents with such decision making capabilities about language processing creates an environment in which it is feasible and useful to attempt even the more sophisticated aspects of language processing in the near term.

1. Introduction

The availability of truly sophisticated, multifunctional intelligent agents is the tantalizing prospect that has been driving work on artificial intelligence for decades. Developing the component capabilities of such agents, however, has most commonly been distributed across different research paradigms. Along with the positive outcomes of this distributed methodology are drawbacks, such as repetition of effort, incompatibility of output formats, and a lack of attention to "boundary phenomena" that transcend individual paradigms. A movement that has usefully stemmed the divide-and-conquer tide is work on cognitive architectures, which offers integrative views of necessary agent functionalities (Langley, Laird, & Rogers, 2009). Work on cognitive architectures resonates with our OntoAgent research program, which is developing intelligent agents endowed with language processing capabilities that are tightly integrated with other functionalities, such as plan-and-goal-oriented reasoning, learning and decision making.

The utility of combining diverse agent capabilities can be seen, e.g., in the Maryland Virtual Patient (MVP) system (Nirenburg, McShane, & Beale, 2008). MVP is a prototype of a clinician training system in which a cohort of virtual patients can be diagnosed and treated by human trainees in open-ended cognitive simulations, with the optional help of a virtual tutor. Virtual patients engage in dialog with a human physician-in-training, answer questions based on their dynamically changing memory, learn new ontological and lexical information, and make decisions about their treatments. Their decisions are affected by their personality traits, physiological and mental states, current knowledge, and goal agenda. The virtual patients in MVP display a large number of the integrated capabilities delineated both by the cognitive architectures community and by the medical community in their quest for virtual patient training systems (Stead & Lin, 2009) – albeit currently for a limited domain. The virtual tutor observes the interaction between the virtual patient and the clinician-in-training and offers contextualized advice.

There is a good reason why we invoke the broad range of agent capabilities in a paper primarily focused on language processing. The topic to be discussed – the resolution of elided scopes of modality – is a complex linguistic phenomenon that, like other manifestations of ellipsis, has essentially remained untreated within natural language processing (NLP) as that field has moved ever farther away from the early AI-NLP goal of deep language understanding. Moreover, even if mainstream NLP did attempt to treat complex phenomena, it is unlikely that the outcome would measure up to the typical definition of success in that community: high rates of precision in evaluating all instances of the phenomenon in a corpus. However, what if the whole problem space were redrawn such that intelligent agents were given the power to determine how confidently they could carry out a *given* instance of language understanding and, in the case of low confidence, decide how to proceed further. For example, an agent could seek immediate clarification of an instance of uncertainty, postpone clarification until and unless it is deemed necessary, or decide not to seek clarification at all. Decisions of this kind involve a whole range of agent capabilities, such as reasoning about the goals and plans (one's own and others'), managing a goal agenda, and determining the extent to which language input must be understood before it can be acted upon. Clearly, language understanding viewed in this perspective is only in part about language; it cannot be viewed as a peripheral input-output problem in cognitive architectures, left to the NLP community to solve in isolation.

Ellipsis is the null realization of a referring expression. Modality is the expression of a speaker attitude[1] – such as *want, hope to, be permitted to, be able to* – that scopes over a proposition. The proposition can be elided if its meaning is readily recoverable from the context. For example, in (1) the scope of the modal element *failed to* is elided (as indicated by [e]), with its meaning being recoverable via a type-coreference relationship with *take action against the perpetrators* from the preceding clause.[2]

(1) Delhi would [take action against the perpetrators] if Islamabad failed to [e].[3]

As a notational aside, within OntoSem, meaning is at the center of all language processing. The meaning of elided elements is determined by examining the meaning of previous utterances, and all meanings are represented using unambiguous ontological concepts, not text strings. However, in order to not overburden the paper with formalism, in the presentation of examples we use square brackets around strings to represent their meanings.

---

1. "Speaker attitude" is a technical term that encompasses such phenomena as polarity, potentiality, volition and other meanings that are not covered by regular dictionary meanings of "attitude".
2. "Type coreference" indicates a different instance of the same kind of event. Here, Delhi's taking action is a different actual event than Islamabad's taking action.
3. Most examples cited here are from the Gigaword corpus (Graff & Cieri, 2003), available at http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2003T05.

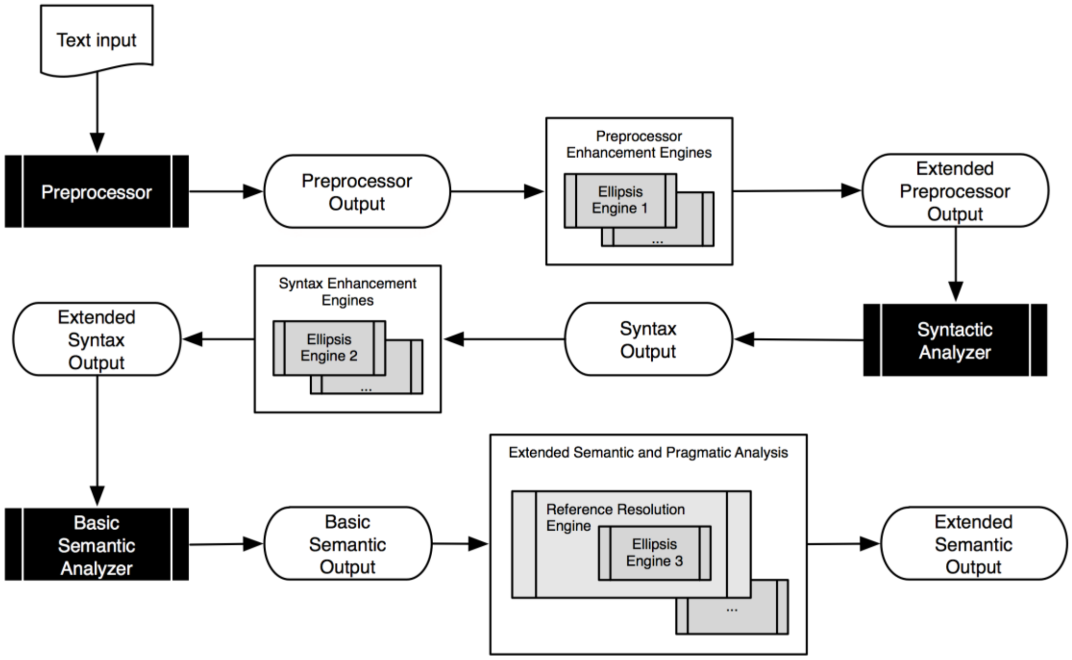

Figure 1. Weaving ellipsis treatment into the language understanding process.

The approach to treating modal scope ellipsis presented here has two innovative aspects with respect to cognitive modeling:

1. *Distribution of the treatment of ellipsis across stages of processing in a psychologically plausible way*. The treatment of modal scope ellipsis is distributed across the traditionally delineated stages of language processing – preprocessing, syntactic analysis and semantic analysis – as shown in Figure 1. We hypothesize that this distribution of effort mirrors certain aspects of language processing by people, such as reasoning by analogy and economy of effort, as discussed further below.
2. *The calculation of expected utility*. We introduce the calculation of expected utility into agent decision making about how rigorously to pursue given instances of linguistic analysis. As mentioned above, it would be unrealistic to require an agent to fully understand every input because such capabilities are well beyond the current state of the art. However, there are two additional reasons why such a requirement is unwarranted: first, many real-world utterances are ill-formed and even ill-conceived, being incomprehensible even to human interlocutors; and second, human interlocutors often happily ignore unclear aspects of utterances that they hypothesize to be unimportant, so why should intelligent agents be deprived of this option? Given these real-world constraints and considerations, a tactically reasonable and psychologically plausible approach is to enable agents to decide when and how deeply to pursue specific language understanding problems in given contexts, using decision functions whose input parameter values can be drawn from any aspect of cognition.

The remainder of the paper is organized as follows. Since the treatment of modal scope ellipsis is motivated by the OntoAgent worldview, Section 2 begins with an overview of this cognitive architecture. Section 3 briefly describes language understanding by OntoAgents, with a special emphasis on the directly relevant microtheories of modality and reference resolution. Section 4 walks readers through the process of detecting and resolving modal scope ellipsis, as illustrated in Figure 1. The discussion includes the practical and cognitively-motivated rationale for each component, available choices for implementation, and how the calculation of expected utility affects language-oriented decision making. Section 5 recaps the main thesis and presents concluding remarks.

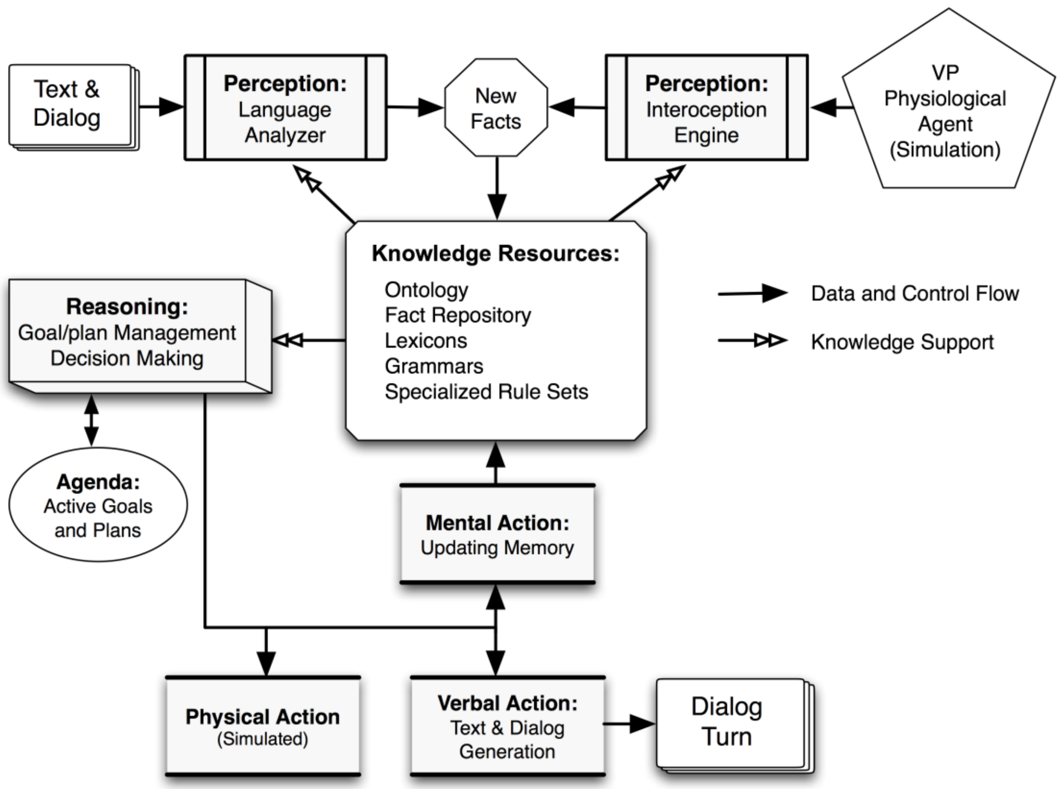

Figure 2. Architecture of agents in OntoAgent.

2. OntoAgents

The OntoAgent cognitive architecture (Figure 2) supports the modeling of human-like behavior in artificial intelligent agents that collaborate with people.[4] The agents in question have simulated bodies and simulated minds, with the latter providing cognitive capabilities

4. For a broader overview of our group's work, see http://www.trulysmartagents.org/index.php. Some aspects of the work presented here are patent pending.

that include interoception (the interpretation of one's bodily signals), learning, planning, decision making, memory management and communication in natural language.

As the figure shows, OntoAgents can undergo two types of perception: *interoception*, which is the experiencing of signals generated by physiological simulation of the agent's body, and *language understanding*, which involves a large battery of pre-semantic and semantic analysis engines. The results of processing input from both modes of perception are formal knowledge structures written in the unambiguous, ontologically grounded metalanguage described by the theory of Ontological Semantics (Nirenburg & Raskin, 2004).

Depending on their content, the knowledge structures are stored in the appropriate knowledge base: ontology, for general world knowledge; fact repository for episodic memories; or lexicon, for newly learned words and phrases. Such structures are the building blocks of agent memory as well as the input to all reasoning processes of the agent (McShane & Nirenburg, 2012). Agent reasoning is carried out at dozens of levels, from the many processes involved in deep natural language understanding, to the processes involved in memory management, to the manipulation of plans and goals. Agent action includes mental actions, like updating memory; verbal actions, like engaging in dialog with a user; and simulated physical actions, like taking medicine or showing up for a doctor's appointment. As mentioned earlier, OntoAgents are at the core of two medically-oriented proof-of-concept systems, Maryland Virtual Patient and Clinician's Advisor (McShane, Nirenburg & Jarrell, 2012).

3. Language Processing by OntoAgents

The purview of Ontological Semantics (Nirenburg & Raskin, 2004), the theory of language processing exploited by OntoAgents, is the automatic semantic analysis of language input. OntoSem, the natural language analyzer that implements Ontological Semantics, processes all texts using the series of processing engines illustrated in Figure 1. OntoSem attempts to generate fully specified, unambiguous ontologically-grounded knowledge structures that are optimized for machine reasoning. (The quality of results of this fully automatic process, naturally, depends upon the domain and complexity of input.) We will briefly illustrate the results of semantic analysis using an example:

(2) I can work now. Under Taliban, I could not [e].

The text meaning representation (TMR) that an agent generates for this input is shown in Table 1, pretty-printed and slightly abridged for presentation. Elements in small caps are ontological concepts. Numerical suffixes indicate concept instances. Each frame is headed by an object or event instance, and its property-value pairs are indented. Each line of the TMR is commented by way of explanation. The property value *find-anchor-time* is a call to a procedural semantic routine that will seek the actual time the text was reported; if successful, the function will return the actual time, which will be used when populating agent memory with this new knowledge. Although details about modal scope ellipsis treatment will be provided in later sections, note at this point that the SCOPE property of the MODALITY-2 frame is currently unfilled, reflecting the ellipsis of the complement of *could not* in the input text.[5]

5. Providing a single example – admittedly slightly too advanced for this stage of the exposition – was deemed preferable to overburdening the text with examples.

Table 1. The TMR for the input **I can work now. Under Taliban, I could not**.

MODALITY-1		
TYPE	POTENTIAL	; *"can"*
VALUE	1	; *highest value on the abstract scale {0,1}*
SCOPE	WORK-ACTIVITY-1	; *"work"*
ATTRIBUTED-TO	HUMAN-1	; *"I"*
TIME	*find-anchor-time*	; *"now"*
MODALITY-2		
TYPE	POTENTIAL	; *"could not"*
VALUE	0	; *negation on scale {0,1}*
SCOPE		; *the ellipsis that must be resolved*
ATTRIBUTED-TO	HUMAN-1	; *"I"*
TIME	< *find-anchor-time*	; *past tense*
TIME	Taliban.TIME	; *"under Taliban"*
WORK-ACTIVITY-1		; *"work"*
AGENT	HUMAN-1	; *"I"*
SCOPE-OF	MODALITY-1	; *an inverse relation*

To give an idea of the size of the OntoAgent language processing environment, the language-independent ontology contains over 9,000 concepts, each of which is described by a large number of properties whose values can be locally defined or inherited. The lexicon of English contains about 35,000 senses, each comprised of linked syntactic and semantic zones, the latter using ontological concepts to describe word meaning. The suite of analyzers has been under development for about 20 years.

The building blocks of Ontological Semantics are *microtheories* devoted to different language phenomena. The microtheories are, at any given time, at different stages of advancement in terms of algorithmic sophistication, coverage, the acquisition of required knowledge resources, implementation, testing and evaluation. There are dozens of microtheories, covering such topics as word sense disambiguation, semantic dependency determination, nominal compounding, treating temporal expressions and processing unknown words. Of particular interest to the current discussion are the microtheories of modality and reference resolution.

Modality is the expression of a speaker attitude that scopes over a proposition. Modality frames in OntoAgent are described by four features, whose value sets are indicated in brackets: **type** {epistemic, belief, obligative, permissive, potential, evaluative, intentional, epiteuctic, effort, and volitive}; **value** {0-1}; **scope** {the meaning of the proposition the modality scopes over}; and **attribution** {by default, the speaker, though third person attribution is possible as well}. The frames for MODALTITY-1 and MODALITY-2 in the TMR above provide examples of the use of these features.

The term "modality" is older and has broader coverage than newer coinages such as "sentiment analysis", "opinion mining" and "subjectivity analysis". Whereas the latter tend to be associated with specific domains (e.g., marketing, politics, national security) and specific methods (e.g., building classifiers in statistically-oriented NLP),[6] "modality" as a topic of study is motivated more by philosophical and linguistic considerations. Fully understanding modal expressions can help an agent to detect the intentions of others, form a profile of others' knowledge and beliefs ("mindreading"; cf. Bello 2011), correctly remember the status of reported events (did/might/should/didn't/etc. happen), and so on.

Modal meanings are detected using the regular language-analysis capabilities of OntoSem – i.e., words and phrases that indicate modality are recorded in the OntoSem lexicon along with their expected syntactic dependencies and their compositional semantics constraints. The OntoSem analyzer uses this information when generating TMRs. Detection of modal scope ellipsis can happen at several stages of processing, but resolution always occurs during semantic analysis, as detailed in Section 4.

As concerns the microtheory of reference resolution, it is quite different from mainstream approaches in current natural language processing. In brief (cf. McShane, 2009 and McShane & Nirenburg, in press, for details and literature reviews):

- rather than treat, as most in the field do, a hand-selected subset of overt referring expressions, we attempt to treat all referring expressions – overt and elided, simple and complex;
- rather than use machine learning trained over a manually annotated corpus, we take a primarily knowledge-rich, rule-based approach;
- rather than considering string-level coreference to be the intended outcome of reference processing, we define reference resolution as grounding new information in the memory of a language processing agent;
- rather than assume that all stages of upstream processing have been carried out to perfection prior to reference processing, we anticipate errors in upstream processing and incorporate them into an agent's decision making about its confidence in resolution results; and
- (of particular importance to this discussion) rather than task an agent with attaining full confidence in every reference decision at any cost, we enable agents to consciously make decisions in this regard.

One aspect of reference treatment is the detection and resolution of elided expressions, and one type of expression that can be elided is the scope of a modal word or phrase. We now turn to the microtheory that treats that latter class of elliptical configurations.

4. The Microtheory of Modal Scope Ellipsis Resolution

To reiterate the core thesis of this paper: Obtaining useful results from ellipsis processing is particularly promising in OntoAgent not only because agents have extensive language processing capabilities but also because they can decide how rigorously to pursue specific instances of language analysis. Such decisions are based on many factors, including their

6. Pang and Lee (2008) provides a nice survey of applications and methods of sentiment analysis, largely focused on clustering methods. They attribute the surge of interest in opinion mining since 2001 to large datasets, machine learning methods, and promises of commercial and intelligence applications.

current goals, their understanding of their interlocutor's goals, their evaluation of how confidently they can carry out resolution, and so on. For example, one OntoAgent (e.g., an agent that tracks the activities of a person of interest) might not be interested in anything counterfactual, thus deciding to ignore all elided scopes of modality in modal frames described as [[type: epistemic], [value: 0]], which indicates negation. Another agent might be tasked to learn as much as possible about its human collaborators, such that all modality frames of [type: belief *or* volitive *or* evaluative] attributed to those collaborators are of high interest and their scopes, if elided, must be resolved even at high cost. Another agent might be tasked to trace what could happen with respect to some object or event in the world, which would give precedence to modal frames of [type: potential] in texts about that object or event. Still another agent might be motivated exclusively by the principle of economy of effort (cf. Cognitive Load Theory; Sweller, Merrienboer, & Paas, 1998), paying little attention to inputs which, on a cursory examination, appear challenging. In fact, we implemented an agent of the latter type for an experiment in resolving modal scope ellipsis in the Gigaword corpus, hypothesizing that an agent could "fill out" at least some of the elliptical gaps in the corpus, thus making the hidden meanings more readily available to other types of NLP engines. Indeed, the agent carried out this task as expected, achieving high precision despite low recall.

Giving agents the ability to choose their linguistic battles models what people seem to do: people do not ask each other for incessant clarifications at every instance of ambiguity or underspecification. We are not, of course, suggesting that our agents, at the outset, will make all of the same decisions as a person: i.e., there will be many cases in which a person would easily make a resolution decision but the agent, unsure, will decide to postpone a decision. We do, however, hypothesize that it will be useful to (1) require that the agent treat the full range of phenomena occurring in natural language interaction, rather than limit its purview to an externally simplified subset, and (2) afford the agent the same decision space as a person when dealing with real, often messy, natural language input. Agents modeled this way will grow in sophistication as a result of two processes: gradual improvement of the microtheories underlying each aspect of language processing and reasoning, and the accumulation of instances of clarification and correction in interactive collaborations with people.

Operationally speaking, the determination of how extensively to process a given input is computed using the second-order features *importance*, *confidence* and *cost*. *Importance* of a particular feature value is calculated using the agent's current plans and goals and its understanding of its interlocutor's plans and goals. For example, if a doctor suggests that a virtual patient have a procedure but the patient doesn't understand what the doctor said about its risks, the patient's goal of actively collaborating in decisions about its treatment plan will advocate asking for clarification in proportion to the importance of this goal to the agent. The *confidence* in a language understanding task is calculated based on the confidence the agent has in each upstream result combined with the confidence of the given resolution algorithm. For example, if the task is disambiguating a polysemous verb, but one of its arguments is an unknown word, then the agent's confidence in the result of disambiguation will, in most cases, be reduced due to the lack of key upstream results to inform the disambiguation task. The *cost* of an analysis task is calculated based on its expected computational complexity, time constraints, demands imposed on a human interlocutor, etc.

The microtheory of modal scope resolution, like all microtheories in OntoAgent, strives to balance psychological plausibility with machine tractability. Consider just a few examples, which anticipate the discussion below. Carrying out ellipsis detection as early as possible both helps to avoid downstream errors by other processors and is in accordance with the psychologically demonstrated human tendency to make decisions earlier rather than later, even in the absence of the full complement of potentially useful information (a well-known decision-making bias is jumping to conclusions, as discussed, e.g., in Kahneman, 2011). Similarly, recording elliptical multi-word constructions in the lexicon both helps the system to correctly analyze such inputs and corresponds with evidence that humans store multiword entities as ready-made units. As Arnon and Snider (2010) report, more-frequent multiword expressions are processed faster by people than less frequent ones. To emphasize, we are not suggesting that our modeling choices reflect in lockstep what people do; however we are suggesting that they follow many of the same principles and thus offer intelligent agents corresponding benefits, such as reduction of cognitive load.

We will attempt to concisely explain the practical implementation of the microtheory of modal scope ellipsis by walking readers through the stages of OntoAgent text processing shown in Figure 1 (a detailed description of text processing in OntoAgent is available in McShane, Nirenburg, & Beale, 2012). The core processors are shown in black. After each core stage of processing, multiple other processors are run, which implement microtheories for the treatment of individual phenomena. Of those processors, we will discuss here only the ones that specifically deal with modal scope ellipsis and directly relevant aspects of broader reference processing.

4.1 Preprocessor → Preprocessor Output

The preprocessor carries out tokenization, part of speech tagging, morphological analysis, lexical lookup, named entity recognition and the recognition of punctuation marks. A subset of this information – specifically, lexemes, their parts of speech, and punctuation marks – is sufficient to permit Ellipsis Engine 1 to detect certain instances of modal scope ellipsis.

4.2 Ellipsis Engine 1 → Extended Preprocessor Output

Ellipsis Engine 1 uses preprocessor output and an inventory of stored patterns to detect certain instances of modal scope ellipsis.[7] For example, the pattern [verbal modal element + period/semi-colon/colon] detects, with high confidence, that example (2) above ends with an instance of ellipsis (could not *do what*?). The engine then inserts a provisional verbal complement into the sentence so that the syntactic analyzer will have one fewer instance of ellipsis to manage – ellipsis being a well-known challenge for syntactic parsers. The inserted verbal complement is supplied with metadata indicating its original elided status so that Ellipsis Engine 2 will know to further pursue the resolution of its meaning.

There are at least two implementation strategies for employing this method of early ellipsis detection: (1) use only extremely high-confidence patterns and accept the ellipsis detection as a final decision; (2) use a broader inventory of patterns (which might result in false positives[8]), associate each with a confidence level (informed by corpus analysis), and submit to the parser both the original input and variant with the hypothesized ellipsis; the parser will then process both variants, submit the results to semantic analysis, and the best overall analysis – which will incorporate the "ellipsis detection" score into its many aspects of scoring – will ultimately be selected.

7. Our approach to making high-confidence, cognitively simple decisions early on is similar in spirit to the sieve approach that has become popular for certain NLP tasks; see, e.g., Ratinov and Roth (2012).

8. We were surprised by how many patterns that we thought would be highly predictive failed to reach a reasonable threshold of confidence upon corpus analysis. For example, the multifunctional comma ended up offering no predictive power since it can be used even to flank adverbs, in which case the complement of a modal can be overt but separate by punctuation: e.g., *He wanted to, in any case, buy a new home.*

The choice of control strategy should, we hypothesize, depend on at least three considerations: first, the cost of introducing multiple variants of an input into an already high-complexity process; second, the degree to which plugging elliptical slots improves parsing as balanced against the degree to which the introduction of incorrectly posited elliptical categories can generate misleading parses; and third, the degree to which ellipsis-detection strategies belonging to later stages of the pipeline can be successful if the initial parse has been confounded by an unrecognized instance of ellipsis.

Ellipsis Engine 1 produces Extended Preprocessor Output, which supplements the original preprocessor output with provisional verbal categories in place of certain elided scopes of modality. Our current implementation strategy is to use only high-confidence patterns and pass to the parser only one preprocessor output for each sentence.

4.3 Syntactic analyzer → Syntax Output

OntoAgent uses the Stanford syntactic dependency parser for the initial stage of syntactic analysis (de Marneffe, MacCartney & Manning, 2006).[9] The aspects of syntactic analysis that are directly relevant for this discussion are clause boundaries, clause ordering, clausal embeddings, and certain syntactic dependencies, described below.[10]

4.4 Ellipsis Engine 2 → Extended Syntax Output

Ellipsis Engine 2 uses the processing results obtained thus far to detect what we call "simple parallel MSE (modal scope ellipsis) configurations", as illustrated by:

(3) He encouraged his children [to take interest in the family business], and they did [e].

(4) Seven golfers, including Leonard, needed to [win] and didn't [e].

(5) They [managed to get out]; his wife did not [e].

(6) I at least wanted to [go three sets] if I could [e].

These configurations contain an elliptical clause directly preceded by a conjunct that is syntactically connected to it in one of several highly constrained ways that can be loosely described as showing syntactic parallelism. Our investigations to date suggest that the clause relationships with the strongest predictive power for modal scope ellipsis resolution are clausal coordination, verb phrase (VP) coordination, parataxis (juxtaposition using certain punctuation marks) and variations on the *if... then* theme (e.g., *if...* [no overt then]...; *if...when*; ... *if...*), as illustrated in turn by the examples above. Such configurations cannot definitively identify the ellipsis sponsor, since that requires semantic analysis as well (see below); however, they so strongly suggest which conjunct contains the sponsor that the agent need look no further. Cognitively speaking, the agent can reduce its cognitive load by accepting this high-confidence solution without launching an unnecessarily complicated search.

The reason for seeking islands of confidence in syntactic parallelism derives from the well-documented effects of parallelism cross-linguistically (see, e.g., Asher, Hardt, & Busquest, 2001; Goodall, 2009; McShane 2005). As concerns ellipsis, it typically imposes a greater

9. Among the responsibilities of the syntax-enhancement engines that we will not discuss here are modifying the Stanford output to make it correlate with the expectations recorded in the OntoAgent lexicon, and reambiguating decisions (such as PP-attachments) that require semantic evidence.
10. Of course, all aspects of syntactic analysis are ultimately important inasmuch as they contribute to semantic analysis, since semantic analysis serves as heuristic evidence for this microtheory as well.

cognitive burden on the interlocutor than an overt category would. In order to fulfill the corresponding discourse obligation, the speaker can foster resolution by employing a highly predictive parallel structure. However, the predictive power of parallel configurations decreases precipitously if the conjuncts – particularly the first – contain relative clauses or other verbal subordinates because such structures provide additional candidate sponsors for the elided verb phrase. For example, if we rewrite example (5) such that the first clause includes several embedded clauses, as in:

(7) They managed to get out because they acted quickly and crossed the border before the troops arrived; his wife did not [e].

it becomes necessary to carry out sophisticated reasoning about the world to determine which action the wife did not do: *arrive? cross the border? act quickly? manage to get out? all of the preceding events together?*

In order to capture the predictive power that syntactic parallelism can provide *by itself*, we introduced the category "simple parallel MSE", defined formally with respect to the output of the Stanford dependency parse. According to the definition implemented in our initial experiments, applicable configurations contained exactly one instance of a CONJ, ADVCL or PARATAXIS dependency, and no instances of CCOMP, PARTMOD, RCMOD, DEP or COMPLM – all of which indicate various types of embedded structures. This definition, while providing almost perfect predictive power, is clearly more restrictive than it would optimally be, since it covers quite a narrow scope of contexts. Surely, syntax should be able to contribute more predictive power, albeit with slightly less confidence, than is harnessed by our definition of "simple parallel MSE".

However, we hypothesize that the best way to further exploit syntactic heuristics – even parallelism-oriented ones – is *not* during this syntax-only early pruning but, rather, in conjunction with semantic heuristics that become available only later on. This hypothesis derives from the cognitively oriented nature of the modeling: at this syntax-only stage of processing, no matter *what* the input means, the agent is making certain ellipsis resolution predictions. As soon as the meaning of the input becomes more important, as it will be for configurations that are syntactically more complex, the associated reasoning should incorporate the of semantic analysis results.

In summary, Ellipsis Engine 2 detects syntactic configurations that confidently predict the conjunct that contains the ellipsis sponsor. This information is recorded for later use by Ellipsis Engine 3, which will determine which *aspects of the meaning of that conjunct* should be used to resolve the ellipsis.

4.5 Basic Semantic Analyzer → Basic Semantic Output

The main goals of *basic* semantic analysis in OntoAgent are lexical disambiguation and the establishment of semantic dependencies.[11] These are represented in basic TMRs of the type illustrated above. A regular part of basic semantic analysis is the generation of both fully specified modality frames, like the one for MODALITY-1 in our sample TMR, and underspecified modality frames, like the one for MODALITY-2. Advanced aspects of semantic and pragmatic reasoning – including things like reference resolution and speech-act detection – are carried out by subsequent engines, resulting in extended TMRs, which are of the same structure but contain additional aspects of meaning.

11. In this discussion, we do not commit to a particular control structure for implementation.

However, certain aspects of modal scope ellipsis *are* treated during basic semantic analysis via lexicalized multi-word expressions that both anticipate the ellipsis and provide information to support resolution. Among the dozens of elliptical multi-word expressions we have recorded in the OntoAgent lexicon are the adverbials *as far as NP can** and *(for) as long as NP (possibly) can**, as illustrated, respectively, by:

(8) Liz Mikropoulos of Bellaire, Ohio [climbed] as far as she could [e].

(9) British actor Daniel Craig, who played James Bond in the latest film about the superspy, said in an interview published Friday he wanted to [continue playing the role] for as long as he could [e].

In the shorthand representations of the expressions, NP refers to a noun phrase of any internal complexity, the parentheses indicate optional elements, and the asterisk indicates any inflectional form of the word.

All multi-word expressions in the OntoAgent lexicon are recorded along with their semantic interpretations: *as far as NP can** adds the property "DISTANCE EFFORT.MAX" to the event in TMR that it modifies, whereas *for as long as NP can** adds the property-value pair "DURATION EFFORT.MAX". So, when the analyzer encounters such an input, it never needs to explicitly recognize that there is an instance of ellipsis: it simply matches the pattern and adds the recorded semantic interpretation to the TMR.

The abovementioned phrases are very frequent, thus justifying their explicit recording. However, they are actually instances of the more general pattern *as ADV as NP can**, where ADV can be any adverb: *as bouncily/precipitously/flamboyantly/... as NP can**. This more generic pattern is also recorded in the lexicon, but its meaning is recorded as a function to be run during the processing of a particular text. The function says, essentially, "Find the property indicated by actual adverb used in the text, make its value EFFORT.MAX, and apply this property-value pair to the meaning of the event to which this modification applies". Functions like these are recorded widely in the OntoAgent lexicon, since many aspects of semantics can be computed only within a specific context. We currently have an inventory of about 25 multi-word expressions devoted to modal scope ellipsis, ranging from very specific to quite generic.

In sum, constructions involving predictable modal scope ellipsis are recorded like all other multi-word expressions in the OntoSem lexicon. Their description includes: (a) the expected syntactic configuration, which can include strings, variables and elided elements; (b) the static semantic interpretation of any required lexical elements; and, if needed, (c) a call to a procedural-semantic routine for resolving the meaning of variable elements. The rationale behind treating frequent collocations as multi-word expressions involves both engineering (improved parsing and semantic analysis) and cognitive modeling (as mentioned earlier, there is evidence that people store frequent collocations explicitly in their mental lexicons).

Since the Basic Semantic Analyzer generates basic TMRs, it provides other types of evidence directly useful for processing modal scope ellipsis. We turn to that evidence now, as it contributes to the work of Ellipsis Engine 3.

4.6 Ellipsis Engine 3 → Extended Semantic Output

The task of this engine is to carry out all outstanding aspects of modal scope ellipsis detection and resolution. At this point in the processing of an input, each instance of modal scope ellipsis falls into one of three categories:

1. The instance has been detected by Ellipsis Engine 1 and Ellipsis Engine 2 has predicted which conjunct contains the sponsor.
2. It has been detected by Ellipsis Engine 1 but, since it does not participate in a "simple parallel MSE" configuration, Ellipsis Engine 2 did not predict which conjunct contains the sponsor.
3. It has not yet been detected.

If the ellipsis has not yet been detected, it can readily be detected in the basic TMR from the empty SCOPE slot in a MODALITY frame. The first task of Ellipsis Engine 3 is to create an inventory of elided modal scopes thus detected. Subsequently, the three categories above become two: either the sponsor-conjunct is known or it is not known. Let us first consider the remaining semantic analysis issues if the sponsor-conjunct is known.

Semantic Analysis Issue 1: Should modal meanings in the sponsor conjunct be included in, or excluded from, the ellipsis resolution? Consider three examples:

(10) The media also blasted Erjavec for taking his wife with him on a trip and insisted he should have [gone through the customs] as all citizens must [e].
[e] = 'go through the customs', *not* 'should go through the customs'

(11) On Friday night, he wanted to [go out in style], and he did [e].
[e] = 'go out in style', *not* 'want to go out in style'

(12) The scheduled train [managed to stop in response to frantic radio warnings], but the supplementary train didn't [e].
[e] = 'manage to stop in response....' *not* 'stop in response...'

In each case, the sponsor conjuncts contain modal meanings scoping over the main proposition; however, whereas the modalities in first two are excluded from the ellipsis resolution, the modality in the last one is included in the resolution, as detailed by the descriptions. The agent decides whether to include or exclude modalities using a rather extensive rule set that seeks to capitalize on generalizations like "an instance of *try to X* is often followed by an instance of *succeed/fail [to X]*". To give a taste of these rules, the salient input parameter values for the rules that cover examples (10)-(12) are summarized in Table 2. The input parameter values include the relative types of modality in the sponsor- and ellipsis-conjuncts, and the correlation (matching or not-matching) between the meanings of the external case-roles in the clause – typically realized as the subject.

Table 2. Sample modality correlation heuristics.

The sponsor clause contains	**The ellipsis-licensing modality is**	**Include in the ellipsis resolution**	**External case-role correlation**	**Ex**
Modality of type M	Modality of the same type, M	Only the scope of the modality	Any	11
Effort or *volitive* modality	*Epiteuctic* or *epistemic* modality	Only the scope of the modality	Any	12
Any type of modality	*Epistemic* modality only (e.g., *did, didn't*)	All modalities (except negation) with the scope	Not matching	13

Semantic Analysis Issue 2: Should other (non-modal) meanings that scope over the main proposition in the sponsor conjunct be included in, or excluded from, the ellipsis resolution? Corpus evidence suggests that several dozen ontological concept types that select event complements tend to not participate in modal scope ellipsis reconstructions. These include DECIDE, CONSIDER, PROMISE, REQUEST-ACTION, ACCEPT, ADVISE, DARE, DENY, INFORM, the first three of which are illustrated by:

(13) Three of the four said they'd decided to [support Bolton], but Voinovich said he could not [e].
[e] = 'support Bolton' *not* 'decide to support Bolton'

(14) For a while, Inka Gawenda said, she thought about [moving back to be close to her family]. But she couldn't [e].
[e] = 'moving back...' *not* 'think about moving back...'

(15) After playing just 20 games last season, he vowed to [return to action this season] but could not [e].
[e] = 'return to action this season' *not* 'vow to return to action this season'

Note, however, that when these are used in correlation with modal meanings that scope over them, ambiguity can arise in the intended ellipsis reconstruction. For example, if (14) were rewritten to include the modal *wanted to,* as shown in (16), the intended ellipsis reconstruction might *in*clude or *ex*clude the non-modal element 'think about'.

(16) For a while, Inka Gawenda said, she wanted to think about moving back to be close to her family. But she couldn't [e].
[e] = 'think about moving back...' *or* 'move back...'

Space does not permit a full discussion of the availability of ambiguity in ellipsis reconstructions, a topic we leave for future reports. Suffice it to say that configurations that predictably permit multiple readings lower an agent's confidence in its selection or one or another reading, as discussed further in Section 4.7.

Semantic Analysis Issue 3. Determine the type of reference relationship between the sponsor and the elided category: type-coreference or instance-coreference. A sponsor and an elided category can refer to the exact same instance of an event (instance-coreference) or to the same type of event but different instances (type-coreference). Typically, if the external case roles – most often, agents – of the events are coreferential, the events show an instance-coreference relationship (as in (14), where the elided move is the same event as the sponsor), whereas if they are not coreferential, the events show a type-coreference relationship (as in (13), where the elided support event is different from its sponsor). The type- vs. instance-coreference distinction – which has not been addressed, to our knowledge, in NLP-oriented studies of ellipsis – is important for memory management in intelligent agents. Figure 1 shows the reference resolution engine that addresses both this issue and the next one.

Semantic Analysis Issue 4: Determine whether internal arguments of the sponsor event should be reconstructed using a strict- or sloppy-identity relationship. Strict and sloppy identity is the argument-level realization of the instance vs. type coreference discussed above (for a theoretical treatment of strict and sloppy identity, see Fiengo and May (1994)). For example, in the example:

(17) Better-off parents could [send their children abroad for English education] but poorer families could not [e].

their children in the sponsor-clause refers to different children than the ones tacitly referred to in the ellipsis clause. The identification of strict vs. sloppy identity of arguments is essential for proper memory management in intelligent agents.

Semantic Analysis Issue 5: Determine whether to include or exclude adverbials from the ellipsis reconstruction. In some cases, like example (17), the adverbials in the sponsor conjunct (abroad; for English education) should be included in the ellipsis reconstruction. In other cases, like (2) above, the conjuncts contain contrasting adverbials (now... Under Taliban), in which case one in the sponsor clause should be excluded from the reconstruction. Our agents currently use a small number of rules – which compare TMR properties like time and location – to make this determination, but a more thorough treatment would increase the agent's confidence in decision making in this area.

This concludes the description of the semantic decisions that an agent must make in order to find the actual ellipsis sponsor in an already-detected sponsor conjunct. If the agent has not yet detected the sponsor conjunct, then the next step depends on when the given instance of ellipsis was detected. If it was only just detected in the TMR, that means that the configuration was not tested for syntactic parallelism by Ellipsis Engine 2. This would occur, for example, if the adverb in sentence (2) were moved to the end of the clause, yielding the input: I can work now. I could not [e] under Taliban. (Recall that Ellipsis Engine 2 uses clause-final punctuation as a necessary detection feature.) In this case, Ellipsis Engine 2 is rerun and, if it predicts the sponsor conjunct, then the battery of Semantic Analysis Issues diagnostics is run, as described in Section 4.6.1 to 4.6.5.[12]

At this point, the only instances of ellipsis that remain to be resolved are those for which no sponsor-conjunct can be confidently detected using strong syntactic heuristics. According to the search strategy currently employed, the agent's next move is to search for a recent conjunct that exploits the modality correlation heuristics described above (and illustrated in Table 1). For example, if the ellipsis is licensed by EPITEUCTIC modality (e.g., *succeed*), the agent will seek a recent conjunct that includes an instance of EFFORT modality (e.g., *try*). The prioritization of modality-oriented heuristics derives from their corpus-attested predictive power even outside of a parallel syntactic configuration. For example:

(18) Brandon said he would like to [find his own lawyer] but was not sure he could [e].

would not be considered a "simple parallel MSE" construction because the ellipsis clause is a subordinate clause embedded in a coordinate clause. However, the modality correlation suggested by the progression *would like to... could,* in conjunction with the proximity of these conjuncts, strongly suggests that the elided scope of *could* should be resolved by the scope of *would like to.*

If the immediately preceding context – currently set to include the given sentence and one

12. We chose not to clutter Figure 1 with arrows indicating the loop from Ellipsis Engine 3, to 2, and back to 3 again.

preceding sentence – does *not* contain any predictive modality correlations, then the agent resorts to a low-confidence method of selecting a candidate sponsor: it walks back through the TMRs generated by text analysis, chooses the most recently encountered event, and evaluates it according to the semantic analysis issues detailed above: e.g., if it is selected by a non-modal special event like the ones listed in 4.6.2, then that event is not included in the ellipsis reconstruction. The precipitous dip in confidence associated with the TMR-walkback strategy actually represents a natural stage in the development of microtheories: for any type of linguistic phenomenon, some instances are readily treated, others are well understood and simply require additional development time, and still others represent a hard residue that will need to be whittled away over time.

4.7 Confidence: Its Evaluation and Consequences

Now let us consider the agent's confidence in various aspects of decision making related to the treatment of modal scope ellipsis and the consequences for its overall functioning. Our glass-box analysis to date suggests the following baseline generalizations: the surfacy detection of modal scope ellipsis by Ellipsis Engine 1 is almost perfect; the detection of "simple parallel MSE" configurations yields almost no false positives but more testing is required to judge the prevalence of false negatives; the detection of ellipsis in recorded multi-word expressions is very good but the coverage of those expressions is not yet optimal; our modality-correlation rules, which are currently supplied with 3 levels of confidence, are on the right track but corpus evidence continues to offer new cases, so we consider this area work in progress; the functions supporting semantic analysis issues 2 to 5 are, similarly, works in progress that fundamentally rely on the correct semantic analysis of the preceding text; detection of elided scopes of modality as empty fillers of the SCOPE slot of a MODALITY frames is quite good, irrespective of the accuracy of other aspects of the syntactic and semantic analysis of the sentence; and finally, as mentioned above, the default, TMR-walkback strategy for selecting an ellipsis sponsor is, as expected, not very reliable.

Most aspects of our modal scope ellipsis processing rely on the output of the Stanford parser and/or the OntoAgent semantic analyzer, both of which are subject to error, particularly for elliptical inputs. Although we have not undertaken to measure the confidence of Stanford parses (syntactic parsing being a capability that we have chosen to import wholesale rather than independently develop), we have begun work on evaluating the system's confidence in semantic analyses. The evaluation metrics involve such features as the number of lexical senses available for each word of input, the extent to which available analyses of arguments fulfill the expectations of the events that select them, the number of words and clauses in the sentence, the depth of embedding in the syntactic parse (with deeper embedding, we hypothesize, suggesting the potential for more errors), and so on.

Earlier, we described how different agents can have different goals that affect the extent to which they pursue difficult language processing tasks. On one end of the spectrum, an agent might be tasked to provide *fast* enhancement of a *very large* corpus for subsequent use by *knowledge-lean NLP engines*. This agent would use only Ellipsis Engines 1 and 2; its coverage would be minimal, its results would be only pointers to the strings that contained the sponsors, but it would work fast and with high confidence. On the other end of the spectrum, an agent might be tasked to collaborate with a person on a highly responsible task, requiring every aspect of text processing shown in Figure 1 to be carried out to a high level of confidence. If such an agent can arrive at a confident *overall text analysis*, then it can act upon that analysis; if not, it will need to clarify whatever remains unclear, from lexical disambiguation to speech act processing to ellipsis resolution.

5. Final Thoughts

This paper has attempted to illustrate the tight integration of NLP with general reasoning in OntoAgents using the example of the microtheory of modal scope ellipsis resolution. We hypothesize that a core reason why many aspects of language are not being treated within mainstream NLP is because they are too difficult to be handled well and in blanket fashion by corpus-based engines given the current state of the art. As Spenader and Hendriks (2005) write in the introduction to the proceedings of a workshop devoted to ellipsis in NLP, "The area of ellipsis resolution and generation has long been neglected in work on natural language processing, and there are few examples of working systems or computational algorithms." In fact, of the ten contributions to that workshop, only one reports an implemented system, the others discussing corpus studies of ellipsis, descriptive analyses of phenomena, or theoretical (typically, pragmatic) frameworks in which ellipsis might be treated. However, when language processing capabilities are incorporated into a multifunctional agent, that agent can make decisions about how deeply to process any given input, giving itself a cognitively well-motivated escape hatch in contexts in which it can judge the given information to be non-critical. This decision making capability, in fact, corresponds with what people seem to do when faced with ambiguous, unclear or contradictory language input.

Acknowledgments

Sincerest thanks to the anonymous reviewers whose insightful comments led to significant improvements not only in the paper, but in the microtheory reported herein. The authors gratefully acknowledge the support of the Defense Advanced Research Projects Agency (DARPA) Machine Reading Program under Air Force Research Laboratory (AFRL) prime Contract No. FA8750-09-C-0181. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of DARPA, AFRL, or the US government.

References

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*, 67-82.

Asher, N., Hardt, D. & Busquest, J. (2001). Discourse parallelism, scope, and ellipsis. *Journal of Semantics*, *18*, 1-16.

Bello, P. (2011). Shared representations of belief and their effects on action selection: A preliminary computational cognitive model. *Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society*.

de Marneffe, M, MacCartney, B., & Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. *Proceedings of the Fifth International Conference on Language Resources and Evaluation.*

Fiengo, R. & May, R. (1994). *Indices and identity.* Cambridge, MA: The MIT Press.

Goodall, G. (2009). *Parallel structures in syntax: Coordination, causatives and restructuring*. Cambridge, UK: Cambridge University Press.

Graff, D., & Cieri, C. (2003). *English gigaword*. Philadelphia: Linguistic Data Consortium.

Kahneman, D. (2011). *Thinking: Fast and slow.* New York: Farrar, Straus and Giroux.

Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, *10*, 141-160.

McShane, M. (2005). *A theory of ellipsis*. Oxford: Oxford University Press.

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

McShane, M., & Nirenburg, S. In press. Use of ontology, lexicon and fact repository for reference resolution in Ontological Semantics. In A. Oltramari, P. Vossen, L. Qin & E. Hovy (Eds.), *New trends of research in ontologies and lexical resources*. Springer.

McShane, M., & Nirenburg, S. (2012). A knowledge representation language for natural language processing, simulation and reasoning. *International Journal of Semantic Computing, 06, 3-23.*

McShane, M., Nirenburg, S., & Jarrell, B. (2012). Modeling decision making biases. *Biologically Inspired Cognitive Architectures.*

McShane, M., Nirenburg, S., & Beale, S. (2012) *Meaning-centric natural language processing*. Unpublished manuscript, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County.

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

Nirenburg, S., McShane, M., & Beale, S. (2008). A simulated physiological/cognitive "double agent". *Proceedings of the AAAI 2008 Fall Symposium on Naturally Inspired Cognitive Architectures*. Arlington, VA: AAAI Press.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1–135.

Ratinov, L., & Roth, D. (2012). Learning-based multi-sieve co-reference resolution with knowledge. *Proceedings of the Conference on Empirical Methods in Natural Language Processing.*

Spenader, J., & Hendriks, P. (2006). Combining multiple information sources for ellipsis. *Research on Language and Computation*, *4*, 327-333.

Stead, W.W. & Lin, H. S. (Eds.) (2009). *Computational technology for effective health care: Immediate steps and strategic directions*. Washington, DC: The National Academies Press.

Sweller, J., van Merrienboer, J. J. G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251-296.