Reference Resolution Challenges for Intelligent Agents: The Need for Knowledge

Marjorie McShane, *University of Maryland Baltimore County*

Natural language processing (NLP) is as yet far from achieving human levels of sophistication. This isn't surprising if we consider that people are amazing processors of language who leverage all of their knowledge of language, the speech context and the world in every language situation.

Treating difficult cases of reference in natural language processing will require intelligent agents that can reason about language and the world using machine-tractable knowledge.

A domain in which the divergence between the abilities of people and the abilities of machines is particularly manifest is reference resolution. Reference resolution is best defined as interpreting the meaning of each referring expression in a language input—like *my finger, JFK, them, ran*—and anchoring it in the mental model (memory) of the intelligent agent processing that input. This semantics-oriented, memory-oriented view of reference resolution is inspired by what people seem to accomplish when resolving reference. It stands in contrast to the more widely pursued NLP task of *co*reference resolution, whose final goal is to match coreferential text strings (words and phrases) with each other, typically with little or no connection to text meaning or memory population and management.

This article provides an example-oriented overview of reference phenomena that are difficult for intelligent agents to process, as well as the types of knowledge, rendered machine tractable, that seem to be required to process them. We begin with a short introduction to "deep semantic" text processing, an approach to NLP that is currently not widely pursued but seems necessary to tackle problems like advanced reference resolution. Next comes an extended example that provides a concrete picture of the problem space in question. Then the full scope of reference phenomena is juxtaposed with the much narrower scope of phenomena that has been treated in systems to date. Comparisons are drawn between the primarily knowledge-lean approach, which has dominated the field so far, and the primarily knowledge-rich approach, which seems necessary for difficult phenomena. Following that are seven high-level questions and their answers that highlight some key challenges faced by reference resolving agents. The article concludes with some thoughts about what to do next in order to make significant progress on reference resolution. The organizational style—example-driven and Q&A—was selected to provide a more

1541-1672/09/$26.00 © 2009 IEEE
Published by the IEEE Computer Society

engaging introduction to the topic than would a formal, linguistically motivated classification.

Semantically-Oriented Text Processing

We can begin to understand the necessary components of and prerequisites for a reference resolving intelligent agent by considering what a person seems to do when reading or hearing a text such as fictional Example 1.

> **Example 1.** *"Next year a film about Mickey Mouse will be released in which he and Minnie go skiing in the Swiss Alps."*

A person's goal is to understand the meaning of the text and either incorporate it into his own knowledge about the world or reject it as untrue. This involves understanding the lexical (dictionary) meanings of the words and phrases in the text, including carrying out all necessary disambiguation; understanding the compositional meanings of word combinations; and incorporating any new information learned from the text into memory. All other aspects of language understanding—such as syntactic parsing and understanding intonation, punctuation, and so on—merely provide intermediate results on the road to the ultimate goal of deriving meaning. During the process of text understanding, a person leverages several types of knowledge, as necessary and available:

- *knowledge of language*, for example, what each of the words and phrases means; how declarative sentences are structured in English; the fact that pronouns like "he" require a textual coreferent unless their meaning can be understood from the extralinguistic context, as by being visible to the speech participants;
- *knowledge about types of objects and events in the world*, for example, what skiing is; the fact that cartoon characters can function like people do (they can ski);
- *knowledge of real-world facts*, for example, what and where the Swiss Alps are; who Mickey Mouse and Minnie are;
- *knowledge of the language situation*, for example, whether the information is coming from a conversation or a newspaper article; how trustworthy the source is; and
- *generalized reasoning capabilities*, for example, that the hearer or reader is expected to know who Mickey Mouse and Minnie are since they are not further described in the text.

As a person builds up an understanding of the input, he determines the role of each referring expression within the larger context.

As a person builds up an understanding of the input, he determines the role of each referring expression within the larger context. If the person already knows about a given entity or event, he can supplement that knowledge with the new information, amend that knowledge with the new information, or reject the new information due to conflicts with what he already knows. If this is a new entity or event, the person creates an "anchor" for it in memory to which the given information is linked, as will be any other information he might learn about the entity or event in the future.

Assume that someone reads Example 1 already knowing who Mickey Mouse is and what the Swiss Alps are, but not knowing about Minnie, the new film, the fact that the film was released, or the event of Mickey Mouse and Minnie skiing in the film. He will

- directly link all new information about *Mickey Mouse* and the *Swiss Alps* to his existing Mickey Mouse and Swiss Alps memories/knowledge;
- create new anchors in memory for the objects *Minnie* and this particular *film*, as well as for the events of *releasing* this particular film and *skiing*, as engaged in by Mickey Mouse and Minnie in the film; and
- establish the text-level coreference relations between *Mickey Mouse* and *he*, and between *the film* and *(in) which*.

This estimation of how a person might understand Example 1 suggests how we might approach enabling artificial agents to do the same.

Let's assume that, like a person, the intelligent agent seeks to *understand* the text, not simply manipulate the strings at a surface level. Let's further assume that, like a person, it has access to the following kinds of resources, whose contents are recorded in a form conducive to machine reasoning.

- An *ontology*, encoded in an unambiguous metalanguage, that contains property-based descriptions of types of entities and events in the world. (The monospace font

`Prestige Elite` is used here to distinguish ontological concepts from words in a language.) The similarity between the names of concepts and English words—for example, the verb *ski* is mapped to the concept `ski-event`—is simply to aid human developers. The concept `ski-event` could also be recorded as A1J90, as far as the system is concerned. The ontology will contain knowledge including that the typical `agent-of ski-event` is `human`, that `cartoon-character` can `function-like human`, and that a typical `location-of ski-event` is `mountain`.

- A *lexicon* that links English words to ontological concepts: for example, the noun *ski* is mapped to the concept `ski` whereas the verb *ski* is mapped to the concept `ski-event`. When the intelligent agent processes texts, it must disambiguate the input words, rendering them in the unambiguous ontological metalanguage, which is more suitable for machine reasoning than are rampantly ambiguous natural languages.
- A *memory* of actual objects and events in the world, which is supplemented as new texts are processed. Entities in memory can formally be distinguished by numbers: for example, `ski-event-4` might refer to Micky Mouse and Minnie going skiing in the film, whereas `ski-event-5` might refer to some real professional skier competing in a race that was reported in the sports section of an on-line newspaper.
- *Rule sets* to support language processing tasks like morphological analysis and syntactic parsing.
- *Processors* that can leverage all of this knowledge, not only for general semantic interpretation of an input—including reference resolution—but also for more generalized reasoning. After all, much of what people glean from language is not stated explicitly.

A formal rendering of the intelligent agent's memories created by reading Example 1 is shown in Figure 1. The formalism is a simplified version of that used in the OntoSem text processing environment, which is the practical implementation of the theory of ontological semantics.[1] This metalanguage was chosen here for illustration for three reasons: first, it has all the expressive power needed for the current analysis of reference resolution; second, it is readable by people; third, agents within the OntoSem environment are already automatically creating such text meaning representations, which suggests that semantically-oriented text processing, although far from a perfected art, is a realistic approach to NLP. There is, however, nothing about this analysis that is specific to a given metalanguage or environment. In the representation, glosses are presented as comments after semicolons. All references have been resolved (Mickey Mouse ~ he; film ~ (in) which; next year ~ 2010).

> Advanced reference resolution will require large, high-quality knowledge resources and reasoners that can exploit them.

Structured, disambiguated text meaning representations like this are more suitable for machine reasoning than raw English text. After all, *next year, he,* and *in which* can't be fully understood until they are disambiguated, contextually grounded and anchored in the agent's world model. For example, if, in 2009, someone asked a question-answering system when the Sydney Olympics occurred, that person wouldn't want the answer, "Next year," extracted from a text written in 1999. Since resolving difficult cases of reference will require complex reasoning by our intelligent agents, and since machine reasoning is more effectively carried out over interpreted than uninterpreted data, teaching the system to semantically analyze language input is well-motivated.

To reiterate the main features of a semantically-oriented approach to reference resolution: First, reference relations aren't established between text strings but between the *meanings* of text strings as represented in an unambiguous semantic interpretation of the text. (In some cases, an utterance might be genuinely ambiguous, underspecified or open to different interpretations. Those issues go beyond the scope of this overview.) Second, the detection of coreference relations within a text, which has been the thrust of most reference resolution efforts to date, isn't the ultimate goal—it is, at most, a step toward the goal of anchoring all referring expressions in an agent's memory and interconnecting them to produce full, contextually coherent memories over which an intelligent agent can reason. Third, advanced reference resolution will require large, high-quality knowledge resources and reasoners that can exploit them. And finally, the complexity of difficult reference issues suggests that a semantically-oriented approach isn't only preferable, it might be inevitable.

```
release-media-12                                        ; will be released
     theme                  film-818                    ; film
     absolute-year          2010                        ; next year
     anchor-time            2009                        ; the year of speech or publication

film-818                                                ; film
     about-as-topic         cartoon-character-10        ; Mickey Mouse
     about-as-topic         ski-event-4                 ; go skiing
     theme-of               release-media-12            ; will be released
     coreferred-with        all-400                     ; (in) which

all-400                                                 ; (in) which—"all" is the root
     corefer                film-818                    ; of the ontology

animate-243                                             ; he
     corefer                cartoon-character-10        ; Mickey Mouse

ski-event-4                                             ; go skiing
     topic-of               film-818                    ; film
     agent                  cartoon-character-10        ; Mickey Mouse
                            cartoon-character-109       ; Minnie
     location               mountain-range-3            ; Swiss Alps

mountain-range-3                                        ; Swiss Alps
     location               nation-13                   ; Switzerland
     location-of            ski-event-4                 ; go skiing

cartoon-character-10                                    ; Mickey Mouse
     agent-of               ski-event-4                 ; go skiing
     has-name               "Mickey Mouse"
     coreferred-with        animate-243                 ; he

cartoon-character-109                                   ; Minnie
     agent-of               ski-event-4                 ; go skiing
     has-name               "Minnie"
```

Figure 1. The semantic representation for Example 1. This formal rendering of the intelligent agent's memories is written in the unambiguous, ontologically-grounded metalanguage used in the OntoSem environment.

A Gallery of Difficult Phenomena

We can get a notion of the range of reference phenomena that next generation intelligent agents will need to treat by analyzing a real-world example such as this:

> **Example 2.** *New advertisements are appearing on Moscow's streets and subways. Comic-book-style stories portray the new quandaries of the Russian middle class. "If we buy the car, we can't afford to remodel the apartment," says a woman with a knitted brow, in one ad. Then comes the happy ending. Her husband replies, smiling: "We can do both! If we don't have enough, we'll take a loan!" ("Buying on Credit Is the Latest Rage in Russia" by Sabrina Tavernise,* New York Times, *January 20, 2003).*

The following analysis focuses on determining which text entities have *textual coreferents* (finding these is the task of most current systems), and which ones should be *directly anchored to memory* (a task pursed by practically no current systems).

- The meaning of *new advertisements, comic-book-style stories, one ad, a loan,* and *a woman* don't require textual coreferents—they're mentions of new entities that must generate new anchors in memory.
- Although *a knitted brow* does not require a coreference link in the text (it, too, is a mention of a new entity)

it must be semantically linked to the woman who's brow it is.

- *Her husband* will also generate a new anchor but it must be linked to the anchor for the woman using a relation indicating marital ties.
- The interpretation of *we* involves combining into a set the anchors for the woman and the man, which do not form a syntactic constituent. Note that the first mention of *we* comes *before* the mention of the woman and the man, thus requiring resolution via *post*cedents rather than the more common *ante*cedents.
- Understanding the meaning of *Moscow's streets and subways* requires linking some unspecified set of streets and subways to the reader's expected knowledge of the Russian city, Moscow. Of course, there are many places called Moscow, so if the human or artificial reader knows of more than one, he or it must select the correct one.
- Although noun phrases with *the* (called definite descriptions) often refer to entities previously mentioned in the text, this is far from always the case—there are conditions under which *the* should not trigger the search for a textual coreferent. Examples include: definite descriptions that have restrictive postmodification (*the new quandaries* is postmodified by *the Russian middle class*, so no previously-mentioned quandaries should be sought); definite descriptions that include a proper name modifier (*the Russian middle class)*; definite descriptions that are clichés and idioms (*the happy ending)*; definite descriptions whose meaning is generic (*the personal check)*; definite descriptions that include semantic ellipsis, that is, the omission of material that is necessary to fully understand the text but isn't syntactically obligatory (*the car* and *the apartment* don't refer to just any car and any apartment—*the car* is the one the couple is thinking of buying and *the apartment* is the one they already own or rent).
- All cases of ellipsis—syntactic and semantic—must be detected and resolved: *If we don't have enough [money]…* Here, the meaning "money" must be understood as what is lacking, and that lacking must be attributed to the couple who needs the money to buy a car and remodel their apartment.
- The meaning of all events—most often realized as verbs in text—must undergo reference resolution since events have just as much referential power as objects. For example, in the following text, the same instance of the event *staying out* is mentioned three times, the third time using the demonstrative pronoun *that*: "Mary *stayed out* late at a party last week. In fact, she *stayed out* well past her curfew and *that* got her grounded for a week."
- *The happy ending*, which must be a new anchor in memory, must be linked to this particular comic-book style story, not to stories in general. In addition, *the happy ending* must be understood as a cliché which, in this context, conveys that the entire passage is a spoof.

> The majority of reference resolution systems to date have treated only those referring expressions that lie on the simpler end of the spectrum.

Clearly, agents capable of advanced reference resolution will need to be able to carry out extensive reasoning about language, the world, aspects of the previous discourse, expectations about what will happen next, and much more if they are to mimic human capabilities of interpreting referring expressions.

Long-Term Needs vs. Current and Near-Term Reality

The previous illustration of difficult cases of reference resolution suggests that artificial agents must be equipped with advanced powers of reasoning and have access to extensive knowledge resources if they are to approach human levels of competence in resolving reference. The problem is that creating high-quality resources and reasoners is long, hard and expensive work. Faced with the choice between declaring a moratorium on reference work until adequate knowledge resources have been compiled or working on reference in a "knowledge-lean" paradigm, the field has opted for the latter. However, this decision has necessitated some compromises. Most notably, the majority of reference resolution systems to date have treated only those referring expressions that lie on the simpler end of the spectrum, and their end goal has been to link text strings to each other. Treating more difficult cases of reference and treating reference in terms of meaning and memory has been considered, thus far, out of purview.

The knowledge-lean paradigm relies on various statistical techniques that are trained over a manually

annotated corpus, typically using a small number of features such as morphological agreement, the text distance between the entity and the potential coreferent, the syntactic positions of the entities (subject, object, and so on), occasionally their semantic roles (agent, theme, and so on), and various other heuristics that do not require text understanding. For background on knowledge-lean reference resolution as well as pointers to the relevant literature see, for example, the work of Ruslan Mitkov[2] and Vincent Ng.[3] Recent work has shown that the incorporation of some semantic features drawn from Wikipedia, WordNet and semantic parsing improves reference resolution for some referring expressions, still within a primarily stochastic paradigm.[4]

While the evaluations of some knowledge-lean systems have been quite good—in the 80th percentile using precision and recall as a yardstick—they must be understood in context. First, the so-called "markables"—that is, those entities that are annotated in a corpus and therefore are included in the purview of reference resolution systems—have been delimited by hand. Second, in most evaluations, the input to reference resolvers has been perfectly preprocessed text—that is, text whose part of speech tagging, syntactic analysis, and so on, have been manually carried out or validated. Third, most experiments have been carried out on grammatically cleaner types of text genres, not, for example, blogs or emails. Mitkov provides an analysis of the extent to which such simplifications of the problem space have inadvertently boosted the impression of the state of the art in reference resolution.[5] The result is that many people, even within the field of NLP, do not realize that such a large portion of the overall problem of reference resolution has not yet been fundamentally addressed.

The knowledge-lean paradigm for reference resolution surged in the 1990s and early 2000s, with progress since being more modest. There's widespread, if not universal, agreement that more knowledge-oriented features are needed to train the statistical systems. Speaking broadly, from the knowledge-lean perspective reference resolution remains difficult because of the following.

There's widespread, if not universal, agreement that more knowledge-oriented features are needed to train the statistical systems.

- Sufficiently large annotated corpora that cover all reference relations don't exist. Manual annotation of corpora is a complex and expensive task typically carried out by undergraduate students. As such, the scope of phenomena covered in annotation projects—at least those that go beyond syntax—is limited, as is the size of annotated corpora. One promoter of knowledge-lean corpus-based methods was the MUC (Message Understanding Conference) reference resolution task, for which sponsors provided annotated corpora for the training and evaluation of the competing systems.[6] Two of the four requirements for the reference annotation strategy were the need for greater than 95 percent interannotator agreement and the ability to annotate quickly and therefore cheaply. These constraints narrowed the scope of phenomena covered by the corpus and, consequently, the scope of phenomena pursued by those utilizing it.
- Not all reference relations can be established at the level of text strings: for example, there can be coreference with an elided category (see my previous work[7]).
- Not all reference relations can be captured using the kinds of syntactic structures output by most parsers: for example, there can be coreference with multiple constituents that do not form a syntactic constituent (see Example 5, p. 55).
- Not all reference relations are *co*reference relations; for example, in the sentence "When *the couple* got home *the husband* started on dinner," *the husband* is part of the set described by *the couple*. This is a reference relation but not a coreference relation.
- There are insufficient features that systems can use for training, such as full semantic interpretations of text entities.

By contrast, reference resolution from a knowledge-rich perspective roughly means that a lot of knowledge—including syntactic, ontological, real-world, semantic and/or discourse-oriented—is incorporated into a reference resolution system. Practically all such systems rely to some degree on statistical support, particularly as a fall-back in situations when knowledge-based methods do not produce any answer at all, or as a tie-breaker in situations when knowledge-based methods fail to select exactly one best answer. Within this paradigm the reasons for narrow

coverage and current suboptimal results include the following:

- insufficient breath and depth of machine-tractable, heuristic-oriented (as opposed to "purely linguistic") descriptions of all reference phenomena;
- the lack of a fundamental understanding of how reference-oriented heuristics, whose values are often competing, can best be combined into resolution algorithms;
- the lack of very large, high quality, broad-coverage, machine-tractable knowledge resources, including lexicons, ontologies and repositories of real-world facts; and
- the lack of processors that can consistently successfully fulfill the prerequisites for high-quality reference resolution, like word sense disambiguation, establishing semantic dependencies among entities in speech/text, and recovering from so-called "unexpected input"—that is, cases in which the agent isn't prepared for some aspect of the input, as by not knowing a given word or word meaning, not being able to parse non-normative syntax, and so on.

Of course the vast majority of practical systems are actually hybrid, meaning that knowledge-based features and statistical methods are used together. The difference between the two roughly circumscribed approaches lies in the comparative weight of knowledge and statistics. Most current hybrids are predominantly statistical. There are, of course, exceptions, systems in which knowledge plays a large or even the primary role. In such environments, all or most of the knowledge-oriented requirements are fulfilled *to the extent possible* considering practical constraints. Such systems typically work best over limited—albeit not only toy—domains. Treating difficult cases of reference will require that the knowledge-oriented contribution of today's systems be significantly expanded.

Seven Questions

This section presents a sampling of some of the more difficult reference phenomena without details (for reasons of space) about how their treatment might be incorporated into the repertoire of intelligent agents. In the spirit of providing glimpses into the phenomena without formal categorization, the discussion is organized around seven questions that specialists and non-specialists alike might find engaging.

> Treating difficult cases of reference will require that the knowledge-oriented contribution of today's systems be significantly expanded.

Which Entities Are Referring Expressions?

This question might, at first glance, seem superfluous: all objects and events that refer to something are referring expressions, and everything else is not. But this intuitive categorization does not readily translate into a sufficient algorithm to permit our intelligent agent to make the referring vs. nonreferring judgment for every text entity it encounters. Consider the following types of entities, which the agent must recognize as *non*referring expressions.

- "It" used without semantic content as a grammatical place holder in certain syntactic constructions: "*It* is raining; *It* is unfortunate that you couldn't come."
- "It" used without semantic content in idiomatic and quasi-idiomatic turns of phrase: "You take *it* for granted that I'll bail you out; Take *it* from me, that will never work."
- Noun phrases used as descriptors rather than referring expressions: "My brother is *a doctor*; That guy is a *regular Don Juan*."
- Metonymic noun phrases that are used as descriptors rather than referring expressions: A man points to the dancers on the stage and whispers to the lady next to him, "My daughter is one of the tutus"—that is, one of the children wearing a tutu.

Clearly, the form of surface strings is an insufficient diagnostic for "referring vs. nonreferring" status, meaning that text understanding must be invoked.

The automatic detection of semantically empty *it* has been worked on with some success (see the work of Richard Evans,[8] among others), but we are still far from achieving error-free detection of all referring vs. nonreferring expressions.

What Does the Referring Expression Mean and Why Does It Matter?

Past work on reference resolution in NLP hasn't relied much on semantics as a heuristic since semantics is difficult and expensive to understand automatically. However, as the examples below show, the surface features and broad generalizations that are used by most current systems can be red herrings, with semantics being the final arbiter for establishing reference relations.

- *There can be gender, number and/or animacy mismatches between coreferents*. For example, *she* (and, more rarely, *he*) can refer to a ship or any other inanimate object one relates to tenderly. Likewise, *they* is making fast inroads, even in writing, as an alternative to gender-neutral *he* or the awkward *he or she, he/she* or *(s)he* ("If someone wants to buy stocks, *they* can do so on the Web").
- *Some referring expressions can have a specific, a generalized or a hybrid referent*. For example, in English the pronoun *you* can refer to one or more specific animate entities ("If *you* eat another piece of pizza *you*'ll explode"); people in general ("It's tough to live in the suburbs if *you* don't drive"); or a hybrid of "you and anyone else in the same position" ("If *you* speed *you* can get a ticket"—assuming the hearer drives a car). See Chapter 7 of my previous work for a related discussion of such phenomena cross-linguistically.[7]
- *Not all definite descriptions (noun phrases with the article "the") require a textual coreferent*, as shown by Example 3. The percentage of definite descriptions that have textual coreferents is actually surprisingly low, having been counted at 50 percent[9] and 37 percent[10] for different corpora. Implemented systems for detecting non-coreferential definite descriptions—that is, noun phrases with *the* that don't require a coreferent—have been developed by David Bean and Ellen Riloff,[10] among others.
- *Text elements can imply meanings that are more specific than what is overtly stated*. There's an indistinct line between a complete semantic interpretation of an input and an interpretation that includes all relevant—or even all possible—inferences. For example, *we* very often encompasses other people not referred to explicitly. In such situations, a system can, on the one hand, represent the meaning of *we* simply as a set in which only some of the members are known. But how much more useful it would be, at least in some applications, if the system would include in its meaning representation the same kinds of inferences that people would make. For example, if a 60 year-old married woman says, "*We're* going to Italy this summer," the implication, unless overridden by contextual clues, is that she and her husband are going; but if she were 30, had kids and was going to Disneyland, the kids would be implied as well. Understanding what, precisely, the woman means by *we* requires knowledge about her and reasoning about the context.

> Viewing text strings superficially cannot provide an intelligent agent with the reference resolving power it needs.

As we can see, viewing text strings superficially cannot provide an intelligent agent with the reference resolving power it needs. For sufficient reference resolution, it really does matter what the referring expressions mean.

Does the Referring Expression Require a Textual Coreferent, an Extratextual Coreferent, or a Direct Link to an Anchor in Memory?

We've already seen cases, such as *he* in Example 1, in which a referring expression requires a textual coreferent. We've also seen cases in which a referring expression can be directly linked to its anchor in memory (*Mickey Mouse* in Example 1). Regarding extratextual reference resolution, the most obvious examples come from the world of embodied conversations, where one can point to something and say "I like *that*," or respond to loud music by yelling, "*That*'s hurting my ears!" In order for an intelligent agent to process such referring expressions, it needs to render its interpretation of real or simulated hearing, vision, smell, taste and haptics using the same or a compatible semantic formalism as is used for the interpretation of text.

What Syntactic Forms Can Coreferential Categories Have?

The most widely treated type of reference resolution carried out by current NLP systems is linking noun phrases with their coreferential noun phrases, as in Example 3.

> **Example 3:** *Jake lent Mollie his sled and watched her speed down the hill.*

However, in many cases, like the following four types of contexts, coreference relations apply between entities that are not noun phrases.

- *Noun phrases can corefer with verbs when both refer to ontological* `events`. Consider Example 4: "The marauders invaded quickly. *The invasion* brought the town to its knees." Given a knowledge-lean

approach, which is primarily syntactic, the noun phrase (NP) *the invasion* in the second sentence must corefer with the verb *invaded* in the first. Permitting cross-part-of-speech coreference significantly increases the search space for coreference relations. However, using a semantic approach, *invaded* is realized as one instance of the ontological event `invade` (`invade-1`) and *invasion* is realized as another instance (`invade-2`). At the level of meaning representations, the coreference between `invade-1` and `invade-2` can readily be established automatically.

- *Noun phrases can corefer with multiple entities that do not form a syntactic constituent*, as shown by the use of *we* in Example 2: *we* refers to *a woman with a knitted brow* and *her husband*, which are presented in different sentences in the text.
- *Referring expressions can refer to an entire span of text*. Some referring expressions—most notably *this, that, it* and various types of abstract nouns, like *argumentation*—can refer to entire propositions or combinations of propositions, as Example 5 illustrates: "The speaker pronounces a 45-minute treatise about the risks of smoking, ending with, 'And *that*'s why you shouldn't smoke!'" The first challenge in treating referring expressions like these is determining whether to seek a single entity as a coreferent (semantically speaking, an `object` or `event`) or one or more propositions. This determination can only be made semantically. For example, in "Give me *that*," *that* must refer to some physical object that can be handed to a person, whereas in "I don't believe *that*," *that* must refer to one or more propositions. In the latter case the question is how many propositions? Research findings suggest that the span of text—or, in semantic terms, the meaning of the span of text—must be contiguous to the sentence containing *this*, *that*, or *it*.[11] But how do we determine how much previous text (or, semantically speaking, how many previous text meaning representations) should be considered part of the coreferent? That is a very challenging problem whose treatment will clearly need to incorporate a significant amount of world knowledge and high-level reasoning.

> There are many kinds of reference relations beyond the commonly pursued "noun phrase to noun phrase coreference."

- *Referring expressions can be elided, as can their coreferents*. When treating ellipsis, the first challenge is to detect that a category is missing to begin with; only after that can its semantic interpretation and reference relations be established. Such detection and interpretation can be supported in a knowledge-based environment by certain types of lexical and ontological knowledge. For example, when one eats, one must eat something, so even though eat can be used without a direct object ("He takes so long to eat!"), the concept it maps to, `ingest`, expects a `theme` that is an `ingestible`. As such, the intelligent agent knows, just as we do, that there is some ingestible involved in the process of eating, should that knowledge be needed for further reasoning.

As we see, there are many kinds of reference relations beyond the commonly pursued "noun phrase to noun phrase coreference," and our intelligent agent must be able resolve them all.

Must Reference-Linked Entities Always be Coreferential?

Not all entities that are linked by a reference relation are coreferential. Instead, they can be in a "set ~ set member" relationship ("I saw *a couple* riding a tandem bike and *the man* wasn't pedaling at all!"), an "instance ~ type" relationship ("I broke *another corkscrew. These things* are so unreliable!"), or they can represent different instances of a given type of entity ("Mom, Jane got *a pony named Jellybean* for her birthday. I want *one* too!"). In addition to these types of non-coreferential reference links, another that has been discussed in the NLP literature is "bridging," defined as a reference relationship by which one entity suggests the existence of another, and that other entity can be referred to as if it had been explicitly introduced into the context. For example, in "Our *swim meet* was cancelled because *the pool* sprung a leak," one can refer to *the pool* using the definite article *the* because people know that pools are expected props in swim meets.

Not surprisingly, descriptive progress on all of these phenomena is well ahead of practical implementations since this type of reference resolution requires both large knowledge bases and powerful reasoning engines.

Is It Always Clear What, Precisely, a Referring Expression Refers to?

The precise referent for a referring expression is not always clear, as

shown by Example 5. Given that the whole speech is about smoking, when the speaker gets to "that's why…," which parts of what he said is he actually referring to? Clearly he's including whatever reasons he gave not to smoke, and clearly he isn't including sidebars about his family or the temperature of the room. But if he presented vignettes about particular smokers' experiences, are those part of the reasons not to smoke? Could a typical listener even remember all of the reasons, or definitively tease apart the reasons from other speech content? Most likely not. This shows that *benign ambiguity* is perfectly acceptable in language and must be allowed for in all theories and implementations of reference resolution (for discussion of multiple possible analyses of referring expressions, see the work of Massimo Poesio and Ron Artstein[14]). The possibility of more than one "correct" answer requires a fundamental redefining of what we consider *sufficient* reference resolution and how we evaluate the capabilities of reference resolvers.

How Can One Determine Whether an Encountered Referring Expression is a New Anchor in Memory or Whether It Should Be Linked to an Anchor Already Present in Memory?

Knowing when to link new information to existing memories can be difficult. Say an intelligent agent encounters the sentence, "Tom Jones came over for dinner"—is this the Tom Jones who is the hero of a Henry Fielding novel, or one of the 10 real people currently listed under Tom Jones in Wikipedia, or someone completely different? Is the listener expected to know who this Tom Jones is, or is the speaker acquainting him with someone new? The movie and the play called Tom Jones can immediately be excluded as coreferential anchors in memory since such entities can't come over for dinner, but what to do next? As always, the trivial answer is, *use the context*—who uttered the sentence, when, to whom, what knowledge those interlocutors share, and so on. For example, if the speaker lives in the 21st century, the baseball player who died in 1923 is out; and if the speaker is your friend, the fictional Tom Jones is out. In short, the more the property values an agent knows about each entity, the better its chances of finding evidence to rule in or rule out candidate reference links in memory. But leveraging that evidence—that is, knowing which property values to compare and when—is the real challenge, especially since an instance of Tom Jones that should create a new anchor in memory might show no conflicting property values with one or more instances of Tom Jones already stored there.

> The possibility of more than one "correct" answer requires a fundamental redefining of what we consider sufficient reference resolution.

This overview of reference resolution has attempted to sketch relevant phenomena with broad strokes, focusing on what might be interesting not only to practitioners of NLP but to anyone who is working toward, or even just hoping for, a really smart, language-enabled intelligent agent.

The overview hasn't been a comprehensive representation of the problem space in that it didn't devote much attention to phenomena at the simpler end of the spectrum. By "simpler" I mean cases like the pronoun *he* coreferring with the only masculine noun phrase in the recent preceding context; the pronoun *I* coreferring with the author of an article; or the noun phrase *the house* coreferring with the only mention of a house in the preceding context.

It is on examples at approximately this level of complexity that current, prevalently empirical, knowledge-lean methods achieve the relatively high evaluation results reported in the literature. Moreover, these results assume perfect preprocessing of the corpus, which is currently beyond the state of the art. However, the effectiveness of such systems drops significantly for inputs in which *he* is used in a context that contains many males; or *I* is used in a quote within a quote; or *the house* is used in a context comparing six different houses. (In this second set of scenarios, we assume that the surrounding context contains plenty of clues to permit disambiguation by the human reader, otherwise the reference resolution task would be moot.) The moral is this: don't use evaluation numbers as a yardstick for the state of the art unless you have read the details of what is being evaluated, and don't assume that continued work in the same paradigms is necessarily the best way to attack the notoriously difficult problem of reference resolution. This is a field in which revolution rather than evolution might well be needed, and in which a radically different conceptualization of the practical constraints—and ways to circumvent

them—might be the key.

As mentioned earlier, attempts are being made to incorporate more knowledge (from WordNet, Wikipedia, hand-created resources, the output of parsers, and so on) into reference resolution systems. Due to the expense of creating knowledge by hand, directions of work that attempt to automatically create knowledge from text—with "knowledge" being understood as information that's more amenable to machine reasoning than raw text—should contribute to the automation of reference resolution. Such directions of work include learning by reading, automatic fact extraction, and automatic lexicon and ontology creation, among others. All of these have been pursued using various methodologies, and their outputs differ widely with respect to the type, scope and quality of knowledge learned. Of course, in linking reference resolution to learning and acquiring knowledge we're simultaneously linking it to the entire field of language understanding and language-oriented machine reasoning. As frustrating as this can be for practitioners of reference resolution—it would be so much nicer if the task were more constrained!—this expansion of the problem space is inevitable. Difficult cases of reference resolution require all the ammunition available to a thinking, reasoning intelligent agent.

Natural questions would be, how often do the difficult reference phenomena described here occur in texts, and do they even deserve our attention? Certainly, one doesn't come across referentially meaty examples like Example 2 every day. On the other hand, Example 2 is a naturally occurring text, not a monstrosity of the type theoretical linguists often use to test the limits of human parsing. In fact, none of the types of phenomena presented here are fringe phenomena. We can't provide actual occurrence statistics because these phenomena haven't been studied widely enough in computational linguistics for such counts to exist. What we would suggest, however, is that if we're looking toward creating intelligent agents capable of human-level performance, they will require the capacity to process all of these phenomena.

So, how should we approach improving the quality of automatic reference resolution engines without having to wait until we've developed a brilliant intelligent agent that embodies the original vision of strong AI?

First, we should continue the work of incorporating into primarily statistical systems semantic features that are automatically extracted from available structured (for example, WordNet, FrameNet) and semi-structured (for example, Wikipedia) sources. Since many such sources weren't created to support NLP (WordNet, for example, was created as a model of a person's psychological organization of the lexicon and only later was leveraged for NLP—with mixed results), they aren't ideal input to machine reasoning. However, they are large and available; and, especially for the near term, learning to better exploit them should lead to improvement in the resolution of some reference phenomena.

Second, we should continue to study difficult phenomena from a linguistic perspective, attempting to develop machine-tractable algorithms that will not only lead to correct reference resolution but will also suggest what kinds of knowledge need to be compiled to support such resolution.

Third, we should continue to manually develop high-quality knowledge resources aimed precisely at supporting machine reasoning, and use bootstrapping techniques to (semi-)automatically expand them over time.

Fourth, we should implement algorithms for resolving difficult reference issues within small domains, for which sufficient knowledge resources can be compiled. Systems that can resolve references over a narrow (I didn't say toy!) domain will not only serve as proof of concept, they can also be expanded over time as more knowledge—generated manually, semi-automatically or fully automatically—becomes available.

THE AUTHOR

Marjorie McShane is a research associate professor in the Department of Computer Science and Electrical Engineering at the University of Maryland Baltimore County. Her research interests include semantically oriented text processing, knowledge representation, and the development of language-endowed, cognitively sophisticated intelligent agents. She received a PhD in Slavic languages and literatures from Princeton University. Contact her at marge@umbc.edu

References

1. S. Nirenburg and V. Raskin, *Ontological Semantics*, MIT Press, 2004.
2. R. Mitkov, "Robust Pronoun Resolution with Limited Knowledge," *Proc. 36th Ann. Meeting Assoc. Computational Linguistics and 17th Int'l Conf. Computational Linguistics* (COLING-ACL 98), vol. 2, Assoc. for Computational Linguistics, 1998, pp. 869–875.
3. V. Ng, "Unsupervised models for coreference resolution," *Proc. 2008 Conference on Empirical Methods in Natural Language Processing* (EMNLP-08), 2008, pp. 640-649.
4. S.P. Ponzetto and M. Strube, "Exploiting Semantic Role Labeling, WordNet and Wikipedia for Coreference Resolution," *Proc. Human Language Technology Conf. North American Chapter Assoc. for Computational Linguistics* (HLT-NAACL 2006), Assoc. for

Computational Linguistics, 2006, pp. 192–199.
5. R. Mitkov, "Outstanding Issues in Anaphora Resolution (Invited Talk)," *Proc. 2nd Int'l Conf. Computational Linguistics and Intelligent Text Processing* (CICLing 01), Springer, 2001, pp. 110–125.
6. L. Hirshman and N. Chinchor, "MUC-7 Coreference Task Definition (Ver. 3.0)," Appendix F, *Proc. 7th Message Understanding Conf.* (MUC-7), Applications Int'l Corp., 1998; www.aclweb.org/anthology/M/M98.
7. M. McShane, *A Theory of Ellipsis*, Oxford Univ. Press, 2005.
8. R. Evans, "Applying Machine Learning toward an Automatic Classification of It," *J. Literary and Linguistic Computing*, vol. 16, no. 1, Oxford Univ. Press, 2001, pp. 45–57.
9. R. Vieira and M. Poesio, "Corpus-Based Processing of Definite Descriptions," *Corpus-Based and Computational Approaches to Discourse Anaphora*, John Benjamins, 1997.
10. D.L. Bean and E. Riloff, "Corpus-Based Identification of Non-Anaphoric Noun Phrases," *Proc. 37th Ann. Meeting Assoc. for Computational Linguistics* (ACL 99), Assoc. for Computational Linguistics, 1999.
11. D. Byron, "Resolving Pronominal Reference to Abstract Entities," doctoral dissertation, tech. report 815, Dept. of Computer Science, Univ. of Rochester, 2004.
12. N. Asher and A. Lascarides, "Bridging," *From Underspecification to Interpretation*, R. van der Sandt, R. Blutner, and M. Bierwisch, eds., *Working Papers of the Inst. for Logic and Linguistics*, IBM Deutschland, 1996.
13. M. Poesio et al., "Learning to Resolve Bridging References," *Proc. 42nd Ann. Meeting Assoc. for Computational Linguistics*, Assoc. for Computational Linguistics, 2004.
14. M. Poesio and R. Artstein, "Annotating (Anaphoric) Ambiguity," *Proc. Corpus Linguistics*, 2005.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.