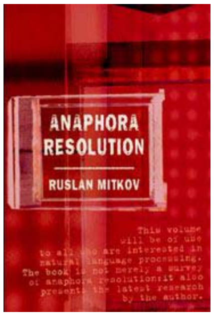
Anaphora Resolution

By Ruslan Mitkov, Longman, 2002, ISBN: 0582325056



Reviewed by Nicolas Nicolov[1]

In the past two decades, as the amount of available information has been growing almost exponentially and data has become ever so plentiful, the gap between existing knowledge resources (in textual, audio and video form) and the ability of computer systems to extract that very knowledge has also been alarmingly widening. The dream of having a piece of wood becoming human, speak and *understand* (Pinocchio) has developed into the idea of having machines not only *speak*, but also autonomously *learn* - that is, acquire knowledge about the world from their language interactions with the environment (as S1mØne, a simulated computer agent, does in a Andrew Niccol's 2002 movie with Al Pacino). The latter aspect of understanding language has been an eagerly anticipated event by early AI researchers, which, by the way, didn't fail to be faced with at times bitter disappointments in those initial years; recently focus has shifted to more tractable concrete problems that allows us to make progress in building models that take advantage of enhanced Natural Language Processing (NLP) capabilities.

Anaphora describes the language phenomenon of referring to a previously mentioned *entity* (also called *object* or *event*); *anaphora resolution* is the process of finding that previous item. Consider the following clarifying example from a British World War II anti-raid leaflet:

"If an incendiary bomb drops next to you, don't loose your head. Put *it* in a bucket and cover *it* with sand."

If this raised eyebrows - don't worry - it is meant to. Indeed "it" could stand for (or *refer* to) either of the two objects mentioned before it, "bomb" and "head". The authors meant the former, but the rules of language have a tendency to bias readers to picking the latter. But then "head"s are not the usual things one puts in buckets and covers with sand. What anaphora resolution, when done correctly, enables us and systems to do, is to merge the previous information about an entity with the new information we encounter. Collecting dispersed pieces of information on an object ultimately builds a fuller picture about it; in technical parlance, systems can store the isolated pieces of knowledge in a knowledge base associated with the *same* object. And the more information we have in the knowledge base, the more new information can be automatically inferred (perhaps using the automated theorem proving technique of resolution as in Horn clause logic). So think of anaphora as the delicate balance between conciseness of communication and the ability of humans to understand each other.

A number of applications, Mitkov says, hinge on systems being able to do anaphora resolution right: machine translation, automatic abstracting, information extraction, question answering. NLP is the arena where the computer scientist meets the linguist; approaches to anaphora resolution require intricate understanding of language phenomena and making them operational requires solid computer science. In the book, thus, Mitkov is addressing a wide audience and illustrating concisely, yet thoroughly, the needed prerequisites. In "Einstein$_i$ felt he$_i$ was on the right track" *he* refers to Einstein (hence we put the same indices *i*). In this case we are lucky there is only one possible item *he* could point back to. Mitkov stresses that more often than not in normal circumstances (texts be them monologues or dialogues) there would be a number of items that could potentially be referred to; this ambiguity gives rise to the need for resolving the correct anaphor (previous item). Although as humans we are amazingly good at picking the right referent (the technical term for the previous item is *antecedent*), for machines this is by far not so straightforward. While the discussion so far might misleadingly suggest it is only pronouns that have the magic property of making us search our mental representations for matching items (pronominal anaphors), Mitkov quickly gives a comprehensive classification of anaphoric phenomena, including "invisible things" (*zero anaphor*) being able to magically refer back. The mystery is due to a peculiarity of Romance languages; in the Italian example, "Judy e' molto intelligente; si e' laureata alla Edinburgh University" ("Judy is very intelligent; [she] is a graduate of Edinburgh University"), Italian allows, actually expects, speakers to drop the pronoun "she" because the morphology of the (reflexive) verb "si e' laureata" [3rd person singular, feminine] makes it clear that Judy is the intended subject in the second sentence.[2]

Conversely languages also allow for pointing back to entities that haven't been mentioned explicitly: "As John was driving, a rabbit jumped on to the road and John slammed on the breaks." "the breaks" refers to the vehicle John is driv-

[1]IBM, nicolas@watson.ibm.com

[2]Incidentally, if you ever wondered why in Italian we can skip the pronouns when the [3rd person singular, present tense] verb provides such strong indications as what the pronoun could be and in English in the same situation we cannot - rest assured in Italian all other pronouns are dropped (hence also the name pro-drop languages). In English the verbs forms for present tense verbs which are not [3rd person singular] are the same and speakers of English would have a harder task of picking the right antecedent.

ing which was not mentioned explicitly.

Mitkov then goes on to describe declaratively what knowledge sources are evolved in the process of anaphora resolution. Consider "The children$_1$ had sweets$_2$. They$_?$ were deli _________." Substituting "delighted" and "delicious" for the last word yields two different antecedents for the pronoun "they", "the children" and "the sweets" respectively (linguists routinely use substitution tests to demonstrate certain constructions are possible and others are kind of odd). The example is a clear case where semantic knowledge about what entities can be delighted and what entities can be delicious helps in picking the right antecedent.

Returning to the applications that need or greatly benefit from anaphora resolution engines, machine translation can take advantage of them, using similar anaphoric expressions in the translation output if the languages are close (e.g., translating from Norwegian into Swedish). Automatic abstracting (summarization), information extraction and question answering all stress the need of being able to piece together knowledge about entities or events which is spread through the information source (not all facts about a person are stated at the point when they are first introduced in the text). Something that Mitkov does not mention (and traditionally not considered a part of anaphora resolution) is that anaphora plays a role in the language generation process. Getting it wrong, as in our World War II leaflet example, brings smiles to people's faces. And in order not to get in wrong technical documentation guidelines recognize the inherent ambiguity of possible antecedents and explicitly try to reduce the chance of the reader getting the antecedent wrong by avoiding certain anaphoric constructions. Incidentally, writing guidelines for lazy readers resort to the same technique - save the reader the effort of finding what you meant by telling him explicitly (perhaps risking a bit of repetition).

Anaphora was appreciated quickly enough as a stumbling block in furthering progress in NLP and a number of theoretical approaches and systems emerged in early 1980s to deal with it. Mitkov does empower the reader with succinct coverage of the theories and the knowledge-intensive techniques of the 80s. AI wasn't "situated" then and researchers would make assumptions about what (pre-)processing was available to them making for elegant theoretical frameworks but not resulting in systems that could easily be applied in practice. Mitkov contrasts this knowledge-intensive approach with later developments that impose fewer requirements on the depths of preprocessing. He calls these techniques knowledge-poor, and as one might guess, these are techniques that derive their "poor" knowledge from corpora. "The pressing need for the development of robust and inexpensive solutions to meet the demands of practical NLP systems encouraged many researchers to move away from extensive domain and linguistic knowledge and to embark instead upon knowledge-poor anaphora resolution strategies." (page 94). An additional factor that enabled less knowledge-intensive approaches to be explored was the availability of both common tools and corpora that permitted the use of machine learning techniques. And finally the field was viewed ripe enough that conferences included tracks on resolving anaphors - the Message Understanding Conferences (MUC6 & MUC7) gave considerable momentum to research in the area. More recently, the Automatic Content Extraction (ACE) evaluation also crucially includes resolving anaphors. Multilinguality is also a factor of concern and researchers are interested in domain- and language-independent techniques. Different languages, though, exhibit subtle differences in the kinds of anaphors they use and their distributions.

Mitkov covers a lot of ground and necessarily at various points needs to refer the reader to the original sources for greater details, though the description of the techniques allows for rational reconstruction of the original work. He does, however, change pace and presents as a comprehensive case study the approach and system he has been developing over the years (MARS). This is the place in the book where the practice of building anaphora resolution engines is fully revealed. The goal is to describe a fully automatic, knowledge-poor, multilingual system. Mitkov does notice a drop of performance when the system works on real output of pre-processing components which are not perfect and make errors; he suggests that previous research should be examined critically in view of many systems having been evaluated under the assumption that they had had access to perfect preprocessing of the input.

The proliferation of approaches and systems begs the question "How do I, as a natural language engineer, choose among alternative anaphora resolution engines?" Corpora with coreference links allow direct comparisons. Mitkov draws a distinction between evaluating an anaphora resolution algorithm and evaluating an anaphora resolution engine as a component of a larger system. For algorithms he presents precision and recall measures, performance measures, comparative evaluation tasks and component measures. For systems Mitkov presents an evaluation workbench where in a plug-and-play mode different engines can be substituted for and the change in performance characteristics observed.

Finally, Mitkov concludes by taking a step back and considering the accomplishments of research in the area of anaphora resolution so far (*Centering* theory about entities in the focus of the attention of the speaker and listener, *Discourse Representation Theory* and how discourse elements are accessed, wide array of systems using different levels of knowledge). He then considers present challenges and directions of future research. Researchers actively working in the area of anaphora resolution as well as graduate students should look here for ways to push the frontiers of science even further.

So are we really close to the moment when S1mØne can understand the questions posed to her without Al Pacino frantically pushing buttons to produce her response? Mitkov says we are 80% there but covering the remaining 20% will not be easy.