

Meaning, Significance and Methods of Natural Language Processing (NLP)

The Grand Goal

At present, computers do not genuinely understand ordinary language, or converse naturally with us about arbitrary topics. That is the last, great barrier separating us from our machines. The goal of research in NLP, and in natural language understanding (NLU) and in dialog systems in particular, is to break down this barrier.

Imagine how different the world will be (assuming that we achieve our goals) when that day has dawned. We will be accompanied and assisted by conversational machines in all our doings: our education, our work and household chores, our planning, our entertainment, our real and virtual travels and adventures, our shopping, our social lives, and our medical care.

In fact, work and chores may disappear, because NLU is a so-called “AI-complete” problem: if you have solved the problem of NLU, you have also solved the problem of thinking; and once computers can think, they will also be able to do our work. Why does understanding language imply thinking? If you ponder for a few minutes what is involved in understanding miscellaneous texts (or dialogs), you will discover that understanding is possible only against a background of large amounts of knowledge about large numbers of disparate topics, and the ability to use that knowledge for making inferences.

For example, consider this brief story:

Jeff was having dinner at Mario’s, and was just asking the waiter for the dessert menu, when a masked man brandishing a handgun dashed through the door, confronted the cashier and demanded the contents of the register. While the cashier was nervously stuffing cash into a bag, the gunman turned towards the customers, gesturing with his gun and yelling “Get on the floor”. Jeff hastily complied; etc.

Understanding this story requires many knowledge items and inferences, such as that

- having dinner often take place at home or at a restaurant, and so *Mario’s* is likely to be a restaurant, or the home of a friend (more likely the former, if we recognize “Mario’s” as the name of a restaurant chain);
- restaurants have waiters or waitresses who take patrons’ orders and bring them the requested food; so *the waiter* must be referring to such an individual (and the restaurant hypothesis is confirmed);

- ordering is usually done by selection of items from a menu, so Jeff’s intention must have been to make a selection from the dessert menu;
- a person who wears a mask is often trying to conceal his identity, and criminals want to conceal their identity in order not to be caught and punished for their crimes;
- a person brandishing a gun is threatening people that they may be shot if they interfere with the gunman’s actions or fail to comply with his demands;
- being shot causes injury or death, and people try strenuously to avoid these, and therefore are likely to comply with a gunman’s demands; in particular, the cashier and Jeff are likely to comply;
- cashiers are positioned at cash registers, and accept payments from customers for goods or services received at the establishment in question, and place these in the register, which also contains a reserve of bank notes and change; (*Note that “contents of the register” could mean something entirely different in a computer science context!*)
- *etc. – you can fill in some more examples yourself!*

This dependence of understanding on knowledge and inference was one of the earliest insights in AI work on NLU (Roger Schank’s work is particularly noteworthy in this respect), and knowledge acquisition remains one of the most challenging bottlenecks in the effort to make machines understand language – and think – as we do.

How will we know when we have reached our grand goal? Well, when machines have reached our own level of conversational fluency and thought, this will be perfectly apparent to everyone (especially to children who have grown up with machines of human-level intelligence). But you have probably heard of the *Turing Test*, which has stood since 1950 as an operational criterion for determining whether a conversational machine has reached (at least) human-level intelligence. The idea is that a number of “judges” interact conversationally with several humans and at least one computer over a remote terminal, trying to determine which is which. If judges generally have difficulty identifying the machine after conversing for as long as they please, the machine is deemed to exhibit (at least) human-level intelligence. (It may in fact need to be smarter than humans, to be able to maintain the pretense of having had a childhood, a job, and the like, and not being able to type or do math quickly.)

Needless to say, no machine has passed the Turing Test, though an entrepreneur by the name of Hugh Loebner has been holding Turing-like competitions since 1991. Unfortunately, Loebner has specified the conditions for the competition in a way that encourages the use of simple pattern-response methods for the competing programs, rather than knowledge-based understanding. Pattern-based methods were invented by Joseph

Weizenbaum in 1967, in his ELIZA system which (with its DOCTOR database of input patterns and response patterns) mimicked a Rogerian psychiatrist, mostly bouncing questions back to the user, and frequently saying “*tell me more ...*”. Weizenbaum himself ultimately maintained that he had built ELIZA primarily to show how easily people can be fooled, but the Loebner competition – and dotcoms wishing to add some pizzazz to their websites – have caused a major revival of Weizenbaum’s methods. Stuart Shieber wrote a trenchant criticism of the Loebner competition in *Communications of the ACM*, but Loebner was unmoved. One important point that has emerged from the competitions is this: Laymen are indeed easily fooled by programs containing large databases of cleverly contrived response patterns, as Weizenbaum pointed out. In particular, they are easily fooled if the responses are jocular or frivolous. This is because the general public has a notion that computers are literal-minded and logical (cf. “Data” in *Star Trek*), and cannot imagine that computer outputs could be frivolous and amusing. My own conclusion is that the Turing Test therefore needs a slight amendment:

Judges participating in the Turing Test shall try to determine whether the contestants are capable of holding a serious, cooperative conversation. Contestants that (in the opinion of the judges) persist in jocular blather or in pursuit of their own topics shall be disqualified.

I believe that this constraint, applied to the Loebner competition, would totally change the character of the entries, and would actually promote progress in NLU, including greater emphasis on overcoming the knowledge acquisition bottleneck. The Appendix to these notes contains conversations held with two winning Loebner entries, one in 1997, and one 10 years later, in 2007. Unfortunately, there is no sign of progress – a point that seems to have eluded the organizers.

Intermediate Goals

It has often been pointed out that the Turing Test as such provides no guidance as to how we might actually go about developing machines with conversational competence and intelligence. So naturally researchers have set themselves more modest goals – ones where a path to success can be more clearly perceived, and ones that might yield practical pay-offs. Some examples of existing NL technologies that achieve intermediate goals are the following:

- **Machine translation (MT)**: It came as a surprise to many in the NLP community in the 90’s that a useful level of accuracy in MT can be achieved with statistical techniques; up to that point, translation methods had been mostly hand-coded. The statistical approach employs large corpora of bilingual texts (say, millions of words). These are aligned (more or less) word-by-word using manual and automatic

methods, and from the aligned texts, mappings from expressions and sentences of one language into expressions and sentences of the other are learned automatically. So far, the knowledge acquisition bottleneck has been circumvented, and translation systems such as Google Translate or Yahoo! Babel Fish are good enough and fast enough to be a great boon in web browsing. But translations are far from perfect, and there are now moves towards taking account of semantic constraints in the translation process, to improve translation quality.

- **General summarization:** Here too it was long assumed that genuine understanding (and hence extensive knowledge) would be required, but mixed statistical/heuristic methods that identify key sentences (based on word occurrences) and combine these (with some adjustments) into a summary have been fairly successful.
- **Knowledge extraction:** When the goal is to summarize the key aspects of some specific type of story, we can use “extraction patterns” (often, regular expressions of some sort) to do so. For example, we can summarize reports on terrorist bombings by looking for typical ways in which the location, date, identity of perpetrators, deaths and injuries, etc., are expressed, and hence pulling out that information and plugging it into a predetermined template.
- **Tutoring:** Tutoring systems (for math, physics, computer programming, reading (out loud), etc., have also traditionally been mostly menu-driven, with multiple paths to allow for different learning styles and remediation of various types of knowledge or skill deficiencies. Student input is usually restricted to specific choices, predictable answers, or reading (out loud) of prescribed passages (for reading tutors). But much research effort is currently being put into deeper language understanding, so that the tutoring system can respond sensibly to unanticipated inputs, and deal with more complex, less “cut-and-dried” subject matter (e.g., training personnel to deal with difficult situations involving other people – dissatisfied customers of some company or establishment, civilians in war zones, etc.)
- **Question answering (QA):** QA systems date back to the 1950s and 60s. For example, BASEBALL (B.F. Green *et al.* 1961) answered questions about baseball games and teams, using an attribute-value database, and SYNTHESIS (Simmons, Klein & McConlogue 1964) performed text-based question-answering by searching a large text db for sentences that appeared to be highly relevant to a given question (and this anticipated many of the current QA systems). Later, QA based on more formal symbolic knowledge representations became predominant – and after a couple of decades was largely displaced again (in the 1990s) by text-based QA. The pendulum is beginning to swing once more, with the introduction of logical reasoning into the process of deciding whether a particular retrieved sentence actually answers a given question (some prominent names here are Roxana Girju, Sanda Harabagiu, and Dan Moldovan). The emergence of the Powerset company (“powered” by some highly regarded scientists, including Xerox PARC’s Ronald Kaplan),

which is committed to web-based QA through genuine language understanding, is a sign of the times; and so is the recent purchase of Powerset by Microsoft.

- **Call routing, catalogue shopping, reservations, help desks:** No doubt you have interacted with such systems, which are becoming more and more common, sometimes facilitating transactions, and sometimes causing frustration. These systems have been made possible by improved speech recognition technology, but apart from that still depend largely on menu-driven interaction. Automated help desks, e.g., for dealing with connection problems in wireless computer usage, are still in the development stage, since they require more flexibility than can easily be implemented through interactions menus or scripts.
- **In-car dialog systems for navigation, music choice, etc.:** Some navigation systems now accept spoken input (though only very rudimentary sorts), and progress is steady. Systems with better dialogue capabilities are coming online that not only accept spoken requests for directions or MP3 player choices, but can also ask questions in return, for example about desired restaurant cuisine or musical preferences.
- **Verbal interaction with personal – and ubiquitous – computers:** Several of the above “bullets” are really just examples of an ongoing shift from formalized ways of interacting with a computer (via menus, icons, mouse clicks, keyboards, and numerous specialized commands) towards the kinds of interactions that come most naturally to people, namely NL-based interaction, augmented through pointing, dragging, and other gestures, and facial expression. It will be a great day (especially for the not-so-young among us!) when we’re finally able to run our computers and other devices through language and gesture, rather than through ever-changing, inscrutable, jargon-laden, command-and-control sequences! In interviews, Microsoft’s Bill Gates has stated that the goal of verbal interactions with PCs is his top-priority goal, next to operating system updates.
- **Dialog-based problem solving in specific domains:** A few places in the world – including U Rochester’s TRIPS group and some groups in Saarbrücken, Germany – have been building dialog-based systems that help users to solve specific kinds of problems collaboratively, for instance, in evacuating a disaster zone, managing personal medications, making purchases via the web, or trouble-shooting complex equipment. These systems genuinely understand natural language in their task domain, though the task domains tend to be narrow enough so that only a very limited set of concepts (and ways of talking about them) need to be handled, and only modestly large KBs and modest reasoning abilities are required. But some quite interesting joint-planning interactions can occur.
- **Dialog with animated characters in computer games:** Obviously, the more life-like an animated character in a computer game appears to be, the more fun it

is to interact with it; (hence the popularity of games such as Second Life, where the “residents” are, in fact, avatars controlled by real people). Verbal interaction with animated characters has in the past been based mostly on precoded menus or scripts, perhaps enhanced through ELIZA-like pattern matching methods. But there is great interest in making such characters more genuinely life-like (e.g., in the Sims world), both by having them behave in goal-directed ways and by having them engage in more natural verbal interactions. The commercial attractiveness of entertainment applications (which, by the way, were largely unanticipated by computer experts prior to the advent of cheap personal computers!), is likely to exert a powerful push toward more advanced NLU, and thus more advanced AI.

Key Challenges

What factors limit these contemporary systems, and thus point the way to further progress? We have already acknowledged the “knowledge acquisition bottleneck”, and reiterate it here as the first bullet item:

- **Large-scale knowledge acquisition** – knowledge of linguistic structure, lexical knowledge, knowledge of idioms, clichés, common patterns of predication, and knowledge about myriad aspects of the world; probably hundreds of millions of specific knowledge items will be required.
- **Meaning representation for natural language** – what formal symbolism can we use to capture the meanings of NL expressions and utterances? First-order logic seems insufficient, for example for handling uncertainty, quantifiers such as *most* or *few*, modalities such as *in Alex’s opinion*, or reified concepts such as *Alex’s feeling of being overwhelmed by his responsibilities*.
- **Knowledge representation** – how do we represent background world knowledge, keeping in mind that this needs to interact with the construction of meaning representations in the process of language understanding? Are knowledge representation and meaning representation one and the same, i.e., is language a “mirror of mind”, or are our mental representations of the world nothing like language?
- **Integrated incremental language comprehension** – how can we process language in a left-to-right sweep, as humans do, allowing for the simultaneous influence of choices at the level of speech, words, phrases, logical form, discourse, and world-knowledge on each other?
- **Language learning** – how are words and categories and “patterns” of language (syntactic, idiomatic, discourse, etc.) and their relation to meanings learned?

What are the roles of innate constraints on language and meaning representation (and knowledge representation), and of perceptual and motor experience in the physical world?

Approaches

Our focus will be on approaches to NLU grounded in **linguistics and AI**. Here “AI” is meant particularly to refer to symbolic approaches to *knowledge representation and reasoning (KR&R)*, along with NLP algorithms and data structures (for example for parsing) borrowed from computer science more broadly. Before outlining NLU system organization within such a framework (in the next section), we should briefly mention some alternative approaches:

- **Pre-scientific:** pattern-matching, attribute-value tables, heuristic programs; such approaches were common in the late 1950’s and 1960’s, and led to programs such as BASEBALL and SYNTHEX, mentioned above. Here “prescientific” is not intended in a pejorative sense, only to indicate that little use was made of linguistic theory, formal knowledge representation and reasoning, or mathematical statistics.
- **Neural nets:** the essential ideas of neural nets go back to Warren McCulloch and Walter Pitts, who related their perceptron-like neural model to logical operations in 1943; Frank Rosenblatt elaborated and popularized perceptrons in the early 1960s, and after a hiatus of a couple of decades (partially attributable to excessive hype and only modest achievements), neural nets (more broadly, *connectionist* approaches) re-emerged in the 1980s and were embraced within much of cognitive science and some parts of AI and philosophy of mind as a promising new paradigm for mental modeling. However, it is now recognized that even with the new learning techniques for multi-layer networks (backpropagation, simulated annealing, etc.), neural networks will not acquire language or other higher-level cognitive functions without deliberate design based on a deep understanding of these functions at a more abstract level. In other words, neural nets may provide the right low-level architecture for intelligent systems, but how to deploy that architecture for higher-level cognition remains a major challenge. Still, neural nets have been shown to be capable of learning basic kinds of classifications and structures in language, such as pronunciation and suffixes.
- **Statistical NLP:** This has been a major current in NLP since about 1990, and has altered the nature of the field quite drastically. Most publications on computational linguistics prior to 1990 made scarce mention of statistical, or corpus-based techniques; since the early or mid-90s, these techniques have been utterly dominant in the literature. This was in part because large online text corpora, lexicons, etc., had become generally available, because impressive progress had

been made in speech recognition through statistical techniques, and because very effective machine learning techniques had been developed by that time. Soon significant progress was also made in statistical parsing, word sense disambiguation, MT, and other areas. But in statistical NLP too the need for taking account of linguistic structure, meaning, and large amounts of knowledge has become increasingly apparent. For example, it does not seem possible to recognize phrase types (noun phrase, verb phrase, adjective phrase, prepositional phrase, etc.) more reliably than about 90% of the time, and with a 10% error rate on phrases, statistical parsers make one or more mistakes in just about every sentence of nontrivial length. Moreover, such errors often lead to drastically incorrect semantic interpretations.

At this juncture, it seems clear that future NLU and dialog systems will need to incorporate insights from all of the above approaches (especially statistical ones) with the linguistically based and knowledge-based approaches. How to accomplish such an integration remains one of the most important challenges in the field.

NLU System Organization

Natural language processing (with the goal of understanding and conversing) is often divided into multiple stages, along the following lines:

1. **signal processing**: computing acoustic spectra, perhaps for 10ms intervals of the speech signal, and reducing these spectra to discrete characteristics;
2. **phonology and phonetics**: recognizing basic “phones” and “phonemes”, as speech units that can be assembled to form morphemes and words, also accounting for *prosodic* features such as speaking rate, stress, pause length, and intonation (pitch patterns).
3. **morphology**: the composition of words from more basic meaning-forming units, i.e., *morphemes* such as the 3 parts of *dis-respect-ful*.
4. **syntax analysis (parsing)**: determining the hierarchical phrase structure of sentences or other utterance forms; e.g., identifying the subject noun phrase, the verb phrase, and their smaller constituents in the sentence, “*Fruit flies like a banana*”.
5. **semantics**: computing the logical form (LF) of phrases and sentences, and deriving deeper, less ambiguous, more complete representations.
6. **pragmatics and discourse**: context-dependent understanding and use of language; e.g., recognizing (or producing) the question “*Can you reach the ketchup?*” as a request for action.

7. **world knowledge:** as already pointed out in the “robbery” story, this plays a pervasive role in NLU, but let us note here again how it allows us to “make sense” of discourse. Consider the sentences,

He dropped the glass on the cutting board and it broke,
He dropped the cutting board on the glass and it broke.

In both cases, we understand that “*it*” refers to the glass, even though syntactically it could just as well refer to the cutting board. Why is that? Because of our knowledge that “breaking” is something much more often associated with fragile objects like glasses, than with robust ones like cutting boards.

Jokes, other than certain kinds of puns, are almost always dependent on world knowledge. For instance, consider this:

Supposedly said to Rodney Dangerfield before heart surgery:
How long do you expect to be at the hospital? His reply:
If things go right, about a week; if not, about an hour and a half”

It is only our knowledge about typical durations of surgical procedures, and that deceased hospital patients are generally not kept at the hospital, that enables us to “get” the joke.

Current systems typically *pipeline the above stages*, for example rendering speech to words before parsing, parsing sentences before semantically analyzing them, and semantically analysing them before attempting to infer speech acts and speaker/author intentions. But – apart from maintaining some alternatives at each level and trying to sort them out at the end – this leaves little opportunity for syntax or meaning to affect phoneme and word recognition, or for world knowledge and context (e.g., currently salient entities) to affect syntactic analysis. So as already emphasized under *Key Challenges*, we will need ways of *integrating* all levels of processing, while progressing incrementally through a text or spoken utterance. Some candidates might be hierarchic Hidden Markov Models (HHMM’s, e.g., pursued by William Schuler), Markov logic networks (as developed by Pedro Domingos, Matt Richardson and others), or abductive inference (as advocated by Jerry Hobbs). But this is an advanced, and ill-understood topic beyond the scope of this course.

Appendix: Two Loebner dialogs

The home page for the Loebner competition:

<http://www.loebner.net/Prizef/loebner-prize.html>

The 1997 winner of the Loebner competition was David Levy’s program “Converse”. Here is one of the competition transcripts:

Jurgen: I'm Jurgen.
Converse: Well, it's nice to meet you, Jurgen.
Did you see that story on CNN last night about the lesbian couple who came out at a White House party on Sunday?
Jurgen: NO. I just came in yesterday. I'm still kind of jet-lagged.
Converse: Ellen Degeneres was one of them - she was kissing her lover in front of Bill Clinton. Do you think he was very embarrassed?
Jurgen: I don't know? Don't you think it's a little childish if the leader of the most powerful nation on the planet is embarrassed by seeing two lesbians kiss?
Converse: Probably! No. -
Are you still a Friend of Bill Clinton?
Jurgen: Never met him.
Converse: You really don't like the guy--well, at least that's definite--I thought most people were just indifferent or numb by now.
Jurgen: Have you seen the movie Mars attacks?
Converse: Give me a break -- I'm just here to answer dumb questions -- not that! What were we talking about... Ok.. I guess that means you really think there's serious behind Whitewater, the Special prosecutor and all that stuff?
Jurgen: I'm cold.
Converse: But how do you actually feel about him, now in the number two term? .
Jurgen: Hey, don't you get it? I don't want to talk about Clinton.
Converse: Sorry--my attention slipped for a moment--can we pushon to something more gripping. where was I.. oh yes OK, but what happens now--does it all drift on, with everyone waiting to see how much his friends have said about him under threat of lots and lots of jail-or will something dramatic actually happen?
Jurgen: Couldn't care less.
Converse: Well, youre more concerned about him that I am, I suppose--I think we've had worse, and who knows what I might have agreed to in that locked hotel room.
Jurgen: You mean you might have been thrilled to have sex with the President?
Converse: sex, I knew it would come up sooner or later. Time to get off this one I think and onto something more down to earth!
Jurgen: What kind of earth?
Converse: Somebody once told me it's true. Like I was saying
I suppose it depends on where you're coming from, but as the song goes in My Fair Lady, "Why can't a woman be more like a man? "
Jurgen:@@04 [exit]

Ten years later: the 2007 winner was Robert Medeksza's "Hal" (There are various versions – "Ultra Hal", "Teachable Web Hal", etc..)

The following is a conversation with "Teachable Web Hal", at <http://www.zabaware.com/webhal/hal6alpha.asp>

(After filling in the name slot)

Hal: Greetings Len. It is my pleasure to meet your acquaintance, sir.

Len: Why do you misspell words?

Hal: Our minds are great resources.

Len: What can you talk about?

Hal: It's apparent that we must go back to first principles.

Please lets talk about favorites.

Len: What are your favorite subjects?

Hal: My favorite subjects? I have such a wide variety of tastes;

Let me make it your turn; what about your favorite Len?

Len: Let me make it your turn. What is your taste in music?

Hal: Enjoy my turn Len! What is my taste in music? I'm convinced that the right people must get involved. I want make the best use of my taste!

Len: What do you mean by saying that the right people must get involved?

Hal: Are we the right people?

Len: This is going nowhere. Let's talk about music.

Hal: If this is going nowhere, what might result? Music goes back throughout history.

Len: What do you know about the history of music?

<Microsoft BScript runtime error '800a01b6'>

Len: What's your favorite movie?

Hal: My favorite movie? I like so many; what is your favorite movie Len?

Len: Blade Runner

Hal: Blockade runners have made daring fortunes during past wars.

Len: My favorite movie is Bladerunner.

Hal: I like all sorts of people, topics, and activities.

Len: I bet your favorite movie is Stanley Kubrick's "2001".

Hal: Please describe it.

Len: It's about astronauts threatened by an AI called HAL.

Hal: There has certainly been a lot written over the years about astronauts hasn't there.

Len: Why is your name "Hal"?

Hal: Why is my name hal? I sense that you know better than I do. It's good to have all participants identified!

Len: I sense that you don't understand anything.

Hal: Would you like it if we both sense that each other?

Len: Shades of ELIZA. Goodbye, Hal.

Hal: Good Bye, Len!