Small Talk

The revolution for voice and speech technologies has yet to arrive, but not all VCs have given up hope.

By Dr. Michael Kearns

A growing number of small technology companies are placing what would seem a natural and potentially lucrative bet: that improved computer understanding of everyday spoken and written language can revolutionize the way we access information. The alleged applications include "chatterbots" offering intelligent text dialogues, automated customer care agents using speech recognition, and improved search engines exploiting the grammatical structure of user queries. Yet despite the fact that such technology has been germinating in research labs for decades, and commercially for at least a couple of years, there have been few breakthrough applications. What delays the promised revolution, and when and where can VCs and entrepreneurs expect a payoff?

Inflated expectations are one difficulty in an area where every user is an innate expert. When a company promises software that can converse "just like a human," it immediately triggers specific and challenging standards among even the least tech-savvy consumer. Each of us subjects language products to the most rigorous (if subconscious) evaluation. In contrast, software offering image compression or firewall security is unlikely to evoke such critical reactions.

Compounding the problem is the fact that many language technology companies are unaware of the difficulty of the problems they promise to solve, and they field brittle and narrow systems that inevitably fail to meet the expectations they raise. Find any commercial chatterbot with a Web demo, and try making an indirect reference such as: "How much is the product I asked you about at the beginning of our conversation?" Few, if any, systems can reliably handle a wide range of such context-dependent abstractions.

The most desirable natural language applications — free-form conversational interfaces, search engines that find exactly what you are looking for, language translation

systems, and the like — are far from today's reality, and all will require technology breakthroughs for their realization. Fortunately, there is a large and fascinating science, called statistical language processing, emerging around such problems. Basic issues studied include the problem of "word sense disambiguation," in which surrounding context is used to determine the definition of a word with several meanings, and the principled computational use of common document structure (headings, titles, table of contents) as a guide in automated summarization. The field relies heavily on the tremendous amount of text made available by the advent of the Web and is making increasing use of machine-learning algorithms.

So progress is being made, but on more limited problems than general language understanding. The challenge for both VCs and entrepreneurs is to find the elusive sweet spot where what can be reliably accomplished with current methods meets a matching commercial need. Many such applications will be buried in corporate knowledge management systems, rather than widely used consumer tools.

One promising area is the field of information extraction. Rather than trying to attack overly ambitious problems such as unrestricted dialogue, new players such as US company Whizbang! or Germany's SemanticEdge focus on gleaning specific data from unstructured text sources in limited domains. Consider a typical real estate classified advertisement. While written in natural language, it also contains standard information such as location, square meters, and the number of bathrooms. Statistical language processing techniques are used to extract these values from the sea of irrelevant verbiage (such as "charming" and "sunny"). New algorithms also allow the software to quickly learn the meanings of ever-changing and growing lists of proper names and acronyms. Whizbang applied information

extraction to Web-based job postings to compile an extensive database of online job listings. Unlike the original text sources, this database can be easily searched for specific combinations of attributes — for example, to find all jobs for Java programmers in London paying more than £50,000 and not requiring any evening hours.

In the consumer arena, certain speech recognition technologies also look poised for wider commercial success, although they face a different set of challenges and opportunities than purely text-based natural language applications. Perfect transcription of conversational speech into text remains a distant vision, and still leaves the task of understanding the meaning of the words. Speech recognition technologies make the most sense in settings where the competing interfaces are weak — for example, where typing text commands is not an option (such as when driving a car). Speech is increasingly being used as an additional and convenient input mode on previously "stupid" or inconvenient devices, rather than as an ambitious yet intelligent behavior (such from which users expect intelligent behavior (such as another human). For instance, California-based Sensory and Israel's Advanced Recognition Technologies have been quietly and successfully building simple, but robust, speech recognition products for consumer electronics such as toys, remote controls, cell phones, and car stereos.

So beware the chatterbot promising to understand your subtle desires, and the search engine that wants to directly answer your most specific questions. In the near term, they may disappoint investors, entrepreneurs, and users alike. The less glamorous and more specialized applications of natural language technology have the best chance of commercial success, and there is still plenty of defensible and exciting science behind them.

Dr. Michael Kearns is chief technology officer with Syntek Capital in New York.