

Our ambiguous world of words

May 31 2013



Words. Credit: jah on flickr

(Phys.org) —Ambiguity in language poses the greatest challenge when it comes to training a computer to understand the written word. Now, new research aims to help computers find meaning.

The verb run has 606 different meanings. It's the largest single entry in the *Oxford English Dictionary*, placing it ahead of set, at 546 meanings.

Although words with multiple meanings give English a linguistic richness, they can also create ambiguity: putting money in the bank could mean depositing it in a financial institution or burying it by the riverside; drawing a gun could mean pulling out a firearm or illustrating a weapon.

We can navigate through this potential confusion because our brain takes into account the context surrounding words and sentences. So, if putting money in the bank occurs in a context that includes words like savings

and investment, we can guess the meaning of the phrase. But, for computers, so-called lexical ambiguity poses a major challenge.

"Ambiguity is the greatest bottleneck to computational knowledge acquisition, the killer problem of all [natural language processing](#)," explained Dr Stephen Clark. "Computers are hopeless at disambiguation – at understanding which of multiple meanings is correct – because they don't have our world knowledge."

Clark leads two large-scale research projects – recently funded by the Engineering and Physical Sciences Research Council and the European Research Council – that hope to overcome this bottleneck. Applications of the research include improved internet searching, [machine translation](#), and automated essay marking and summarisation.

"Many of the recent successes in language processing such as online [translation tools](#) are based on statistical models that 'learn' the relationship between words in different languages. But if we want the computer to really understand text, a new way of processing language is needed," said Clark.

As [Eric Schmidt](#), Executive Chairman of [Google](#), said in 2009: "Wouldn't it be nice if Google understood the meaning of your phrase rather than just the words that are in that phrase?"

Clark has turned to quantum mechanics and a longstanding collaboration with Bob Coecke, Professor of Quantum Foundations, Logics and Structures at the University of Oxford, and Dr Mehrnoosh Sadrzadeh, Queen Mary (University of London), who works on the applications of logic to computer science and linguistics.

"It turns out that there are interesting links between quantum physics, quantum computing and linguistics," said Clark. "The high-level maths

that Bob was using to describe quantum mechanics, which also applied to some areas of computer science, was surprisingly similar to the maths that I and Mehrnoosh were using to describe the grammatical structure of sentences.

"In the same way that quantum mechanics seeks to explain what happens when two quantum entities combine, Mehrnoosh and I wanted to understand what happens to the meaning of a phrase or sentence when two words or phrases combine."

Until now, two main approaches have been taken by computer scientists to model the meaning of language. The first is based on the principle in philosophy that the meaning of a phrase can be determined from the meanings of its parts and how those parts are combined. For example, even if you have never heard the sentence the anteater sleeps, you know what it means because you know the meaning of anteater and the meaning of sleeps, and crucially you know how to put the two meanings together.

"This compositional approach addresses a fundamental problem in linguistics – how it is that humans are able to generate an unlimited number of sentences using a limited vocabulary," said Clark. "We would like computers to have a similar capacity to humans."

The second, more recent, 'distributional' approach focuses on the meanings of the words themselves, and the principle that meanings of words can be worked out by considering the contexts in which words appear in text. "We build up a geometric space, or a cloud, in which the meanings of words sit. Their position in the cloud is determined by the sorts of words you find in their context. So, if you were to do this for dog and cat, you would see many of the same words in the cloud – pet, vet, food – because dog and cat often occur in similar contexts."

Working with researchers at the Universities of Edinburgh, Oxford, Sussex and York, Clark plans to exploit the strengths of the two approaches through a single mathematical model: "The compositional approach is concerned with how meanings combine, but has little to say about the individual meanings of words; the distributional approach is concerned with word meanings, but has little to say about how those meanings combine."

By drawing on the mathematics of [quantum mechanics](#), the researchers now have a framework for how these approaches can be combined; the aim over the next five years is to develop this to the stage that a computer can use. Clark has spent the past decade developing a sophisticated parser – a program that takes a sentence of English and works out what the grammatical relationships are between the [words](#). The next step is to add meaning to the grammar.

"To solve disambiguation and build meaning representations of phrases and sentences that computers can use, you need lots of semantic and world knowledge. The idea is to take the parser and combine it with the word clouds to provide a new meaning representation that has never been available to a computer before, which will help solve the ambiguity problem.

"The claim is that language technology based on 'shallow' approaches is reaching its performance limit, and the next generation of language technology will require a more sophisticated model of meaning. In the longer term, the aim is to introduce additional modalities into the meaning representation, so that computers can extract [meaning](#) from images, for example, as well as text. It's ambitious but we hope that our innovative way of tackling the problem will finally help computers to understand our ambiguous world."

Provided by University of Cambridge

Citation: Our ambiguous world of words (2013, May 31) retrieved 20 March 2024 from <https://phys.org/news/2013-05-ambiguous-world-words.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.