Retelling the story of Babel

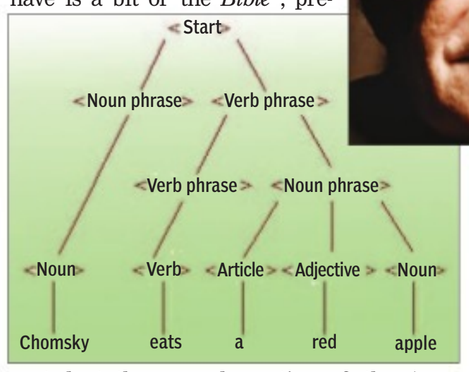
THE BIBLE HAS STEPPED IN TO HELP TRANSLATE FROM AND TO LESSER LANGUAGES, WRITES **S ANANTHANARAYAN**

The Old Testament has it that soon after the Flood, a united people who spoke in one language were fast building a tower, the one called Babel, that would reach high into the sky. God wished to stop the progress and He sent down different languages. This cut communications among the workers and the project came to a standstill. Language since has long been the barrier in human enterprise.

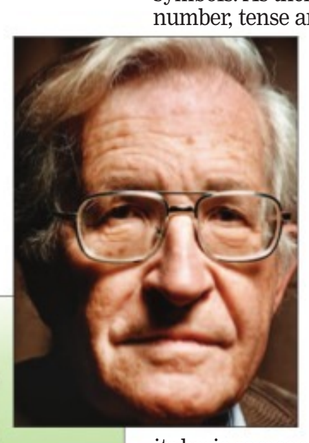
Technology, with software that readily translates text from one language to another, seems at last to be connecting different tongues. On the Internet and even cell phones, this happens almost seamlessly, so that persons using different languages do not even notice that the other person is using another. The best translation programmes, however, need large resources of specifically prepared text matter in the different languages to be able to bring this about.

These resources are available with the major languages like English, French, German, Spanish, Chinese and also a great many others, but not with lesser used languages like Faroese, spoken in the Faroe Islands, between Norway and Iceland, or Gallican, spoken by a community in north-western Spain, or Akawaio, a Caribbean language, Aukan spoken by a group in Surinam or Cakchiquel, a Mayan dialect, or even some Indian languages. Technology is, hence, not able to connect speakers of these languages with the vast academic and commercial material now on tap in other languages.

A team working in the University of Copenhagen, however, hopes to set this right with the help of the *Bible*, which contains sizeable text that has been translated into almost every language. Professor Anders Søgaard, with Seljko Agic and Dirk Hovy, at the Centre for Language Technology at the university, have described their work, which uses existing translations of the *Bible* as a bridge to help lesser languages access the translation resources available for the major ones in their paper, "If all you have is a bit of the *Bible*", presented at the annual meeting of the Association of Computational Linguistics, which was held at Beijing in July this year.



The special preparation of the text matter, which translation software needs, consists basically of adding a label to identify words according to the role, like the part of speech that the words play in the sentence. A human bilingue goes about translation by understanding meaning from text in one language and then expressing the same message in the second language. A machine process, on the other hand, has no understanding and only manipulates symbols. As there are different rules of gender, number, tense and word order in different languages, a simple word-by-word substitution would not work and there has to be an intermediate, abstract form derived from the text in one language before it can be converted into text in the second language.



Chomsky hierarchy

Important work on analysis of language, to understand how sentences are formed, was carried out by language philosopher Noam Chomsky. The Chomsky hierarchy breaks up a sentence in any language into its basic components, like nouns, adjectives, articles or verbs, arranged according to the rules that apply to that language. An instance is shown in the illustration of the analysis, from bottom up, of the sentence, "Chomsky eats a red apple". The sentence can be seen as deriving from the basic form of a <noun phrase> and a <verb phrase>.

The <noun phrase> reduces to just the proper noun "Chomsky". The <verb phrase>, however, has structure; it first reduces to a <verb phrase> and a <noun phrase>. The second order <verb phrase> goes to the verb "eats", while the <noun phrase> has more structure, of an <article>, "a", an <adjective>, "red" and the <noun> "apple".

Now, with this starting structure of <noun phrase> and <verb phrase>, we can attempt expressing the sentence in French. The first <noun>, which is a singular, masculine noun, goes just to "Chomsky". In the <verb phrase>, the verb "eats" translates to "mange", which takes care that it is in the singular, and not "mangent", which is plural, if there had been two persons going at that apple. The next <noun phrase> has three components, like this: first is "une", which is the article "a", in the singular, and feminine, as "apple" in French is feminine, the second is "rouge", for "red", which, in this case, does not change form for masculine or feminine, and then "pomme" for apple. But the rule in French is that the adjective goes after the noun that it qualifies. The sentence hence becomes "Chomsky mange une pomme rouge", with no thought for who Chomsky is or the colour of apples — but we do need two inputs — the equivalent words for "Chomsky", "eats", "a", "red" and "apple" and also that they are nouns, verb, article and adjective, so that we can apply the proper rules.

Data base

While we could find bare equivalent words in a dictionary, this would not serve the purpose in most cases, as many words have alternate meanings and are also used as different parts of speech in different contexts. For a computer system to be able to recognise these differences, what we need is mass of existing text matter where, ideally, all the words that we need have been used, in different contexts, and a label has been attached to each word to show the nature of the use of the words, that is to say, as what part of speech, like noun or verb, and also its context.

Attaching such labels to words in a list is called "tagging" and a collection of text that has the classifying data attached is called a "tagged corpus". If such resources exist, they enable computers to rapidly extract the structure of text in one language and generate text in another language. Software builds on this basis with more analysis, like statistical data, and the result is very efficient translation, "on the fly".

But the basis for the software to function is the existence of classified text collections that helps identify individual words and their role in a sentence or a group of sentences.

In the major languages, these resources have been in creation since decades, but not with minor languages or those not in the mainstream. This is where the University of Copenhagen group has made use of the *Bible*, which consists of parallel text in all languages. As the text of the *Bible* follows rigorous translation, verse by verse, "word alignment" of the versions in different languages leads to acquisition of "parts of speech taggers" for the lesser language based on the existing tagging of a brace of major languages. The *Bible*, hence, serves as a bridge for tagging resources to become available across languages.

"The *Bible* has been translated into more than 1,500 languages… and the translations are extremely conservative, the verses have a completely uniform structure… we teach the machines to register what is translated and find similarities between the annotated and unannotated texts so that we can produce exact computer models of 100 different languages," says Søgaard. The resources have been made available to other developers and researchers, he says, which should help create technology for languages such as Swahili, Wolof and Xhosa that are spoken in Nigeria!

THE WRITER CAN BE CONTACTED AT
simplescience@gmail.com



Anders Søgaard, Dirk Hovy and Seljko Agic

MITIGATING TOXICITY

TAPAN KUMAR MAITRA LOOKS INTO THE REGULATIONS PUT IN PLACE BY THE UNION GOVERNMENT FOR PESTICIDE CONTROL

The toxicity of pesticides to humans, their ability to remain in the environment and accumulate in products require the establishment of strict scientifically substantiated regulations for their safe application. In India, the rules for using pesticides are worked out together by the Union ministries for agriculture and health. Every year, an approved "List of Chemical and Biological Means for Controlling Pests, Plant Diseases and Weeds Allowed to be used in Agriculture" is jointly issued by the ministries. Consequently, when employing pesticides, it is necessary to adhere to the list approved for the current year and also abide by the instructions on the application of the pesticides compiled in strict conformity with the requirements adopted for the relevant substances.

The recommended rates of usage must be observed strictly. Excessive use of a particular pesticide may lead to its large accumulation in the environment and in the products obtained. To protect the health of the population and prevent the circulation of pesticides in nature, sanitary norms have been established for the maximum tolerated concentrations of pesticides in the air of the working zone, water of open basins and in the soil. Of special significance is control of the amount of pesticides in the soil because it constitutes the source of contamination of food products, water basins and air. The content in the soil is controlled in the spring before field work starts. When the topsoil contains persistent pesticides (carbaryl, HCH, polychloropinene, toxaphene) in amounts exceeding the maximum tolerated levels, only grain and technical crops may be grown and surface treatment with these pesticides is prohibited.

For the sanitary control of pesticide residues in products, standard values of the tolerance levels in various food products and fodder are determined for each substance. These indices are coordinated by the United Nations Food and Agricultural Organisation and the World Health Organisation. The values of the maximum tolerated residual amounts (tolerance levels) are established from the results of tests after studying the toxicity of a pesticide to animals and from the determination of the dynamics of the residues in the relevant culture. Cultures differ substantially in their ability to accumulate pesticides. For example, with a toxaphene content in the soil of 10 mg/kg, plants absorbed from two to 46 mg of the pesticide per kilogram of their mass.

The smallest values of the tolerance levels are established for dietetic food products (milk, eggs, meat) and somewhat higher ones for vegetables and fruit. For example, in India the tolerance level of HCH (a mixture of hexachl-orocyclohexane isomers) in meat, eggs, milk, dairy products and sugar must never exceed 0.005 mg/kg, in cereal crops 0.2 mg/kg, and in potatoes and vegetables 0.5 mg/kg. No residues of heptachlor, carbaryl, 2, 4-D, mercury compounds, polychloropinene, parathion-methyl, thiram, etc, are tolerated in any food products.

The magnitudes of the tolerance levels vary with progress in investigations of the toxicity of pesticides and their influence on the environment. These levels are used to establish the harvest time (the time that must elapse after the last treatment with a pesticide before a crop may be harvested). All pesticides applied to plants or incorporated into the soil not later than harvest time remain in amounts not exceeding the tolerance levels or completely decompose. It is determined by the persistence of the substance and also by its toxicological and sanitary properties. For rapidly detoxicating substances, harvest time is two to 20 days, and for more toxic ones a couple of months.

When persistent pesticides have to be used, nevertheless, within shorter periods or on crops in which they may accumulate, regulations for using the products obtained are established. To prevent the possible poisoning of workers in fields treated with pesticides, the periods when work is permitted (re-entry times) and the conditions of work are regulated. For example, after the use of persistent pesticides such as HCH, work not associated with soil cultivation may be performed after six days, while work associated with cultivation of the soil may be performed only after two weeks.

Another important condition for ensuring the safety of workers and protection of the environment from contamination with pesticides is the strict observance of the safety rules in the storage, transportation, and use of toxicants in agriculture released by the Union ministry for agriculture and approved by the Union ministry for health. Strict observance of the standards and rules adopted for controlling pesticide use will ensure reliable prophylaxis of poisoning and effective protection of the environment.

THE WRITER IS ASSOCIATE PROFESSOR, HEAD, DEPARTMENT OF BOTANY, ANANDA MOHAN COLLEGE, KOLKATA, AND ALSO FELLOW, BOTANICAL SOCIETY OF BENGAL, AND CAN BE CONTACTED AT tapanmaitra59@yahoo.co.in

Getting 'under the bonnet'

SCIENTISTS CLAIM TO HAVE FOUND A CORRELATION BETWEEN HOW WELL WIRED-UP SOME INDIVIDUALS WERE TO THEIR COGENITIVE ABILITIES AND GENERAL SUCCESS IN LIFE. **STEVE CONNOR** REPORTS

The brains of high-achieving individuals are wired up differently to those of people with fewer intellectual or social abilities, according to one of the first studies to find a physical link between what goes in the brain and a person's overall lifestyle. An analysis of the "connectivity" between different parts of the brain in hundreds of healthy people found a correlation between how well wired-up some individuals were to their cognitive abilities and general success in life, scientists said.

The researchers found that "positive" abilities, such as good vocabulary, memory, life satisfaction, income and years of education, were linked significantly with a greater connectivity between regions of the brain associated with higher cognition. This was in contrast to the significantly lower brain connectivity of people who scored high in "negative" traits such a drug abuse, anger, rule-breaking and poor sleep quality, they said.

"We've tried to see how we can relate what we see in the brain to the behavioural skills we can measure in different people. In doing this, we hope to able to understand what goes on 'under the bonnet' of the brain," said Professor Stephen Smith of Oxford University, who led the study published in *Nature Neuroscience*.

The scientists were part of the $30-million Human Connectome Project funded by the US National Institutes of Health to study the neural pathways of the brain. Connectomes have been likened to taking real-time images of the living circuit diagrams governing the communication of signals from one part of the brain to another. They compared the "connectomes" of 461 health people taken by real-time brain scanners called functional Magnetic Resonance Imaging and attempted to see if there were any significant correlations with 280 different behavioural or demographic measures, such as language vocabulary, education and even income.

White matter fiber architecture from the Connectome Scanner dataset.

Each fMRI analysis looked at the connectivity — the amount of nerve signalling — that takes place between about 200 different regions of the brain. The one that stood out was the connectivity between the parts of the brain involved in so-called higher-level cognition, such as language and learning, Professor Smith said. "You can think of it as a population-average map of 200 regions across the brain that are functionally distinct from one another. Then we looked at how much all of those regions communicated with each other, in every participant," he said.

"The quality of the imaging data is really unprecedented. Not only is the number of subjects we get to study large, but the resolution of the fMRI data is way ahead of previous large datasets."

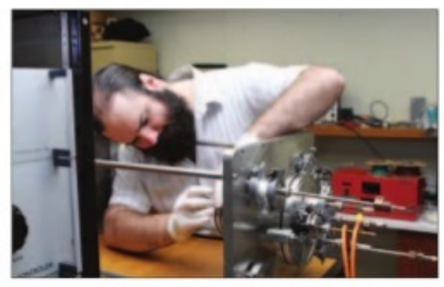
The ability to measure the amount of nerve signalling between different parts of the brain, especially those involved in high cognition such as learning and memory, could help scientists to better understand the nature of general intelligence, which is currently measured by tests that examine a range of intellectual skills.

"It may be that with hundreds of different brain circuits, the tests that are used to measure cognitive ability actually make use of different sets of overlapping circuits," Smith said. "We hope that by looking at brain-imaging data we'll be able to relate connections in the brain to the specific measures, and work out what these kinds of test actually require the brain to do."

It may also be possible to use the research to work out how to train people to improve their brain connectivity and therefore push than up the scale so that they achieve more than they otherwise would, he added. "It's a question of whether it's possible to move people up the axis of connectivity. We know from other research that it is possible to improve cognitive performance with training, but what we don't know yet is whether this is true of connectivity," he explained.

THE INDEPENDENT

PLUS POINTS



One tank is all

Paddy Neumann, a PhD student at the University of Sydney claims to have created a technique that has the potential to take humans to Mars and back on a single tank of fuel. He says his ion thrusters work by hurling particles backwards so that a spacecraft can be propelled forwards, the *Daily Mail* reported.

The ion drive works by bombarding the fuel source with electric arcs that cause ions to be discarded. These ions then move through a magnetic nozzle, resulting in forward propulsion. The innovative part of Neumann's drive is the type of fuel that was used. While the High Power Electric Propulsion system runs on xenon gas, his ion drive can, instead, run on various metals, many of which can be found in space junk.

When it comes to acceleration, High Power Electric Propulsion may not be ideal to launch a spacecraft off a planet. The current record holder for fuel efficiency of an ion drive is the National Aeronautics and Space Administration with its High Power Electric Propulsion system that allows 9,600 seconds of impulse. This is a measure of thruster efficiency and is sometimes called "bounce per ounce".

The new drive developed by Neumann has achieved up to 14,690, according to student newspaper *Honi Soit*. He said it could power a spacecraft to "Mars and back on one tank of fuel" and has has applied for a patent and will be presenting his results at the 15th Australian Space Research Conference today.

Microbes & wine

The distinct regional conditions, or *terroir*, in which grapes are grown are thought to shape a wine's character. But strict scientific evidence of this phenomenon has been lacking. Now, researchers in Auckland, New Zealand, have confirmed that at least one aspect of *terroir* — local differences in yeast strains — does indeed alter the outcome of Sauvignon Blanc fermentation. Their findings were published on 24 September 24 in *Scientific Reports*.

Grape "must" — the freshly crushed fruit, seeds, skins, and stems of harvested vines — can be turned into wine either by inoculating with a pure yeast culture or by allowing the microbes naturally present on the fruit to carry out spontaneous fermentation. It is well known that these naturally occurring microbes exhibit regional variations, said Matthew Goddard of the University of Auckland and the University of Lincoln in the UK. But it was unclear whether these differences translated to distinct wine flavors and aromas, he said. "This study attacks that question." His Auckland-based team first identified and isolated the six major related strains of the yeast *Saccharomyces cerevisiae* present at each of six New Zealand wine-producing regions. They then inoculated these 36 strains individually into sterilised Sauvignon Blanc grape juice to begin fermentation. "We controlled for absolutely everything else other than these microbes," he said, "and then we asked: What are the wines that result from those strains?"

The wines were tested for the presence and quantity of 39 volatile compounds and other standard quality parameters such as ethanol, acidity and sugar. The researchers found that the chemical signatures of the wines showed a large degree of overlap, but that the profiles of those produced by yeast strains from the same region tended to cluster — indicating the wines did, indeed, exhibit a regional character.

RUTH WILLIAMS/THE SCIENTIST

Bumblebee tongues

According to team of researchers led by Candace Galen of the University of Missouri, bumblebee tongues are getting shorter — and it's in part because of climate change, they concluded in a study published on 25 September in ***Science***. In North America's Rocky Mountains, summers have warmed up and flower populations have declined, giving bees with shorter tongues an advantage — they can access nectar from many types of flowers, whereas specialised bees with longer tongues are limited to flowers with a specific shape.