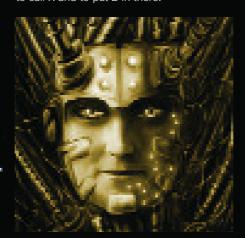# TLHINGAN HOL DAJATLH'A' [DO YOU SPEAK KLINGON]

Language is something we take for granted. We learn it as a child, maybe study a new language at school, or pick up some choice phrases for a holiday trip. A language consists of words - the bits it is made of - and grammar - the way these bits are put together. "The cat sat on the mat" makes sense. "Cat mat on sat the" doesn't. We have the same words but the rules of grammar are broken ... the information is lost. This confused list of words still follows a rule: the words are arranged alphabetically. But without the right rules, the rules of grammar, the words can't do what they need to do, which is let us know where that pesky cat is.

When we have to communicate with a computer and give it instructions we want to make life easy for ourselves. In the beginning programmers were forced to use binary – lists of 1's and 0's to tell the computer what to do. That's because computers are really just a very complex box of switches, and the 1's and 0's told the computer which parts of its circuits to switch on and off, but it made writing the programs a nightmare. It was asking you to speak in an alien tongue.

### A BETTER WAY TO COMMUNICATE WITH OUR COMPUTERS

Something had to change, so those long-suffering programmers looked to what they knew best, human languages, to find a solution. In the same way as a word is made up from letters you could start to think of lists of binary instructions that you could associate together into a simple command to which you could attach some meaningful name. So rather than tell the computer in 1's and 0's ('0101010' say) to put a particular set of values into a particular circuit, you could use the phrase 'LET X=2' to mean the same. The computer understands how to turn a number into binary. It knows that 2 is 10 in binary (that means take a two and add no ones to it, just like a normal decimal 10 means take a ten and add no ones), so this useful idea of a computer language that WE understand means we don't need to talk the nasty alien computer lingo. The language instruction tells the computer to find a place in its memory, which the programmer wants to call X and to put 2 in there.



### SO TO PROGRAMMING LANGUAGES

The programmer's language is translated to the language the computer uses by following a list of instructions. In the same way as with a human language, if we get the grammar wrong and say typed '=LET 2 X' the computer wouldn't know what to do, the information is there but the grammar is wonky. So computer languages developed following many of the rules of human languages to try and keep it as easy as possible for human programmers to program. BASIC, Beginners All Symbolic Instruction Code was an early attempt. That's why, for example, there are university courses on Computer Science and Linguistics. The two are very closely related.

### THE UNIVERSAL TRANSLATOR?

There are now computer programs that translate human languages. You may have seen these on the Google or Babel Fish web sites for example. What is happening here is that, to translate say French text into German, the French is first translated to a language the computer understands and can work with. The text in that computer language is then translated back into German ... so it's passed through that strange alien binary language of the computer somewhere in between. Of course its very basic, Google just uses a simple approach to translation, but there are many researchers around the world trying to develop software that takes into account the complex rules of human languages and grammar. Some systems even try to learn languages from scratch like a child, in effect trying to produce the universal translator similar to the idea in Star Trek, which can translate any language to English. (Of course it would be easier to use the TARDIS telepathic circuits like Dr Who ... a small summer project there for anyone interested:-)

We tend to think of human languages as always having been around in their current form, and computer languages as new, but human languages need to have developed from somewhere, and that's another fascinating story where computers have helped. Some people have even created their own spoken languages such as Esperanto. The basic rules and words of Esperanto were proposed by LL Zamenhof at the end of the 19th century. The idea was to create a world language that everyone could speak. It never really took off, though there are still many people who learn it and use it.

Which brings us to Klingon, a made up language from Star Trek developed by Marc Okrand, (see the Klingon Language Institute at www.kli.org/). It has words, a fixed grammar and there are lots of 'native speakers' out there, so you guessed it ... there are programs that translate English to Klingon. It's interesting to think that a software program to translate a made up alien language does it by using its own made up alien language in the computer. As Mr Spock would say 'Quite Fascinating!'

Qapla' ("Success!")

**Footnote**
In the interests of Galactic Peace it should also be pointed out that there are also programs that translate English into Vulcan, Romulan, Ewok, Wookie, Ferengi ...

Traditionally machine translation has involved professional human linguists manually writing lots of translation rules for the machines to follow. Recently there have been great advances in what is known as statistical machine translation where the machine learns the translations rules automatically.

It does this using a 'parallel corpus': just lots of pairs of sentences; one a sentence in the original language, the other its translation. Parallel corpora are extracted from multi-lingual news sources like the BBC web site where professional human translators have done the translations.

معد ... بالم ، : ذترت بكلة الانيء العراقية ترسية ان ناب رئيس مجلس قيادة الثورة في العراق عزة ابرهيم التقى اليوم الاربعاء في بغداد رئيس مجلس ادارة المركز السعودي لتطوير الصادرات عبد الرحمن الزامل

## Let's look at an example translation:

Above is some original arabic with its translation below:

Machine Translation: Baghdad 1-1 (AFP) – The official Iraqi news agency reported that the Chinese vice-president of the Revolutionary Command Council in Iraq, Izzat Ibrahim, met today in Baghdad, chairman of the Saudi Export Development Center, Abdel Rahman al-Zamil.

Human Translation: Baghdad 1-1 (AFP) – Iraq's official news agency reported that the Deputy Chairman of the Iraqi Revolutionary Command Council, Izzet Ibrahim, today met with Abdul Rahman al-Zamil, Managing Director of the Saudi Center for Export Development.

This example shows a sentence from an Arabic newspaper then its translation by the Queen Mary, University of London's statistical machine translator, and finally a translation by a professional human translator. The statistical translation does allow a reader to get a rough understanding of the original Arabic sentence. There are several mistakes, though. Mistranslating the "Managing Director" of the export development center as its "chairman" is perhaps not too much of a problem. Mistranslating "Deputy Chairman" as the "Chinese vice-president" is very bad. That kind of mistranslation could easily lead to problems. That reminds me of the point in *The Hitch-Hiker's Guide to the Galaxy* where Arthur Dent's words "I seem to be having tremendous difficulty with my lifestyle," slip down a wormhole in space-time to be overheard by the Vl'hurg commander across a conference table. Unfortunately this was understood in the Vl'hurg tongue as the most dreadful insult imaginable, resulting in them waging terrible war for centuries.…

For now the human's are still the best translators but the machines are learning from them fast!