

Scientists discover oldest words in the English language

... and predict which ones are likely to disappear in the future

London - 26 Feb 2009: Scientists at the University of Reading have discovered that 'I', 'we', 'who' and the numbers '1', '2' and '3' are amongst the oldest words, not only in English, but across all Indo-European languages. What's more words like 'squeeze', 'guts', 'stick', 'throw' and 'dirty' look like they are heading for history's dustbin – along with a host of others.

Evolutionary language scientists from the University of Reading, one of the world's leading centres in this field of research, have been investigating how languages evolve, and whether that evolution followed any rules. Until recently they believed they would not be able to track words back in time for more than 5,000 years, however their new IBM (NYSE: IBM) supercomputer has enabled them to go back almost 30,000 years, and finally provide the answers.

The scientists have been able to analyse the family of Indo-European languages – of which English is a modern-day example – reconstruct the rate at which words evolve and predict future changes to our vocabulary. The oldest words we use today have been in existence for at least 10,000 years.

Looking to the future, the less frequently certain words are used, the more likely they are to be replaced. Other simple rules have been uncovered - numerals evolve the slowest, then nouns, then verbs, then adjectives. Conjunctions and prepositions such as: 'and', 'or', 'but' and 'on', 'over', 'against' evolve the fastest, some as much as 100 times faster than numerals. 'Throw' which is expected to evolve quickly, has a half-life of 900 years, there are 42 unrelated sounds for it across all the languages. In 10,000 years time, it will likely have been replaced in 10 of them – possibly including English, unless of course we all do our part to keep the word in circulation.

"50% of the words we use today would be unrecognisable to our ancestors living 2,500 years ago. If a time-traveller came to us, and told us he wanted to go back to that period, we could arm him with the appropriate phrase book, and hopefully keep him out of trouble" explained Mark Pagel, Professor of Evolutionary Biology at the University of Reading.

The IBM supercomputer at the University of Reading, known as ThamesBlue, is now one year old. Before it arrived, it took an average of six weeks to perform a computational task such as comparing two sets of words in different languages, now these same tasks can be executed in a few hours.

Professor Vassil Alexandrov, the University's leading expert on computational science and director of the University's ACET Centre¹ said "The new IBM supercomputer has allowed Reading to push to the forefront of the research community, it underpins other important research at the university, including the development of accurate predictive models for environmental use. Based on weather patterns and the amounts of pollutant in the atmosphere, our scientists have been able to pinpoint likely country-by-country environmental impacts, such as the affect airborne chemicals will have on future crop yields and cross-border pollution".

Caroline Isaac, Deep Computing Executive at IBM said "Supercomputers are enabling the world to become

increasingly interconnected, instrumented and intelligent. We have now reached a tipping point in price/performance that's allowing breakthroughs in university research that were previously unimaginable”.

Notes to editors

¹ACET - Advanced Computing and Emerging Technologies Centre

The Indo-European languages are most of those originally found across Europe, the Middle-East and the Indian subcontinent. Examples include: Celtic, Roman, Greek, Germanic, Nordic (with the exception of Finnish), Slavic, Armenian, Iranian, Afghan, Gujarati, Hindi, Bengali, Nepali and Kashmiri, and of course modern-day derivations such as English and Spanish.

Researchers call words that persist relatively untouched across the ages ‘cognates,’ which means that the words have a systematic sound correspondence that proves their common ancestry. For example, cognates meaning “water” exist in English (water), German (wasser), Swedish (vaten) and Gothic (wato) – read them again and you can discern the ‘aht’ sound common to all. The most resilient cognates, the numerals, have not changed significantly in their entire history.

The half-life of a word is the expected amount of time for there to be a 50% chance for that word to be replaced by an entirely different word.

The research has shown that word types evolve in the following order (from slowest to fastest): numerals, pronouns, nouns, verbs, adjectives, prepositions and conjunctions.

When the IBM supercomputer was installed it was one of the most powerful in the country. It consists of a JS21 cluster, comprised of 700 servers, with a total of 2800 Power PC processors running at 2.5GHz and delivering a peak performance of 28TFlops (28 million million) floating point operations per second.

About the University of Reading

The University of Reading is ranked as one of the UK’s top research-intensive universities. The quality and diversity of the University's research and teaching is recognised internationally as one of the top 200 universities in the world.

The University is home to more than 50 research centres, many of which are recognised as international centres of excellence such as agriculture, biological and physical sciences, European histories and cultures, and meteorology.

The University takes a real-world perspective to its research and is consistently one of the most popular higher education choices in the UK.

For further information visit: www.reading.ac.uk

About IBM

For more information about High Performance Computing from IBM, please visit: www.ibm.com/deepcomputing

<https://uk.newsroom.ibm.com/2009-02-26-Scientists-discover-oldest-words-in-the-English-language>