nature

nature scienceupdate

updated at midnight GMT today is tuesday, july 9

Small word network

English words are connected by just three degrees of separation.
2 July 2002

PHILIP BALL

Close reading: words are interlinked like people are.
© GettyImages

Word association can link just about any two common words in the English language using an average of three steps, says a team of scientists in Arizona.

The semantic links between English words make the thesaurus a 'small world', much as the network of human social interactions connect us all by six degrees of separation, find Adilson Motter and colleagues at Arizona State University in Tempe[1]. The researchers expect languages other than English to have the same properties, even if their syntactic structure is very different.

The researchers traced the links between 30,000 English words in an online thesaurus. For example, the word 'actor' can be connected to 'universe' through two intermediaries. The

relatedstories

*Friends are stranger than strangers
20 November 2001

*Networking can be hazardous
9 March 2001

*Science is all about networking
12 January 2001

*The art of networking
25 October 2000

*Social shortcuts
9 August 2000

*Better networking
24 August 2000

morenews

*Worm turns fish antisocial
9 July 2002

*Patterns shape ant cemeteries
9 July 2002

*Frog sings like a bird
9 July 2002

*Rhinos buddy up
8 July 2002

*New spin on

'universe' through two intermediaries. The thesaurus lists 'character' as a synonym for 'actor'; 'character' is also equated with 'nature'; and 'nature' with 'universe'.

Moving from 'actor' to 'universe' in the network of words therefore takes three steps. To the surprise of Motter and colleagues, they found that the same was true of just about any randomly chosen pair of words in the thesaurus. The English language, in other words, enjoys just three degrees of separation.

Obviously this isn't true for all words - 'vanadium' and 'ambidextrous', say, might be more difficult to connect. But the researchers included in their network only 'root' words, which themselves warrant a primary entry in the thesaurus. Many of the synonyms listed for a given root word are not themselves root words - they don't appear elsewhere in the thesaurus. Root words tend to be the most common or generic - 'big', say, rather than 'gargantuan'.

A small-world network isn't defined solely by the property any two nodes are connected by a small number of steps. Nodes must also be highly clustered - that is, if two nodes are both linked to a third, there must be a high probability that the two are also directly linked. Friendship networks are highly clustered - if you and I share a common friend, we probably know each other too.

Motter's team show that the language network has a high degree of clustering. For example, 'nature' is listed under the entries for both 'universe' and 'world', and 'world' is also listed under 'universe' (and vice versa).

The researchers think that the network structure of a language probably has its origins in the nature of cognition and memory. It is not surprising that language is highly clustered, as we remember things associatively - by grouping similar concepts together. This makes it easier for us to retrieve related words from our memory.

The small-world network also means that

transistors
5 July 2002

*National flock screened for scrapie
5 July 2002

apparently quite different concepts, such as ’actor’ and ’universe’, are closely linked by a short series of semantic steps. This, say the researchers, makes it easier for us to carry out mental searches when using language - we can get to our intended destination quickly, regardless of our starting point. A database cross-referenced in this way would be relatively easy to search computationally.

According to this argument, language has evolved its current structure because it is easy to use. Words acquire multiple meanings because that makes the network highly interconnected, so that the average path length between two words is small. It remains to be seen whether searching our memories for a particular word really entails wandering mentally along links in the network.

References

1. Motter, A. E., de Moura, A. P. S., Lai, Y.-C. & Dasgupta, P. Topology of the conceptual network of language. *Physical Review E*, 65, 065102, (2002).