New research reveals how we make sense of compound words
8 October 2019, by Katie Willis

People process compound words—like snowball—and words that look like compound words but aren’t—like carpet—in the same way, according to new University of Alberta research that has broad applications from rehabilitation after stroke or brain injury to developing AI that understands how humans use language.

“Our results show that when we encounter what looks like a compound word, we can’t help but parse out the constituent parts and then put the word back together—even when it doesn’t make sense to do so,” said U of A cognitive psychologist Christina Gagné, who was a co-author on the research.

“All of this processing happens unconsciously. Your brain sees the word ‘car’ and the word ‘pet’ in ‘carpet,’ even though it is not the most efficient way of processing the word.”

The researchers found that people had a harder time figuring out when “pseudo-compound” words like carpet were misspelled than when actual compound words like snowball were jumbled into misspellings like "snobwall."

“The reason for that is that there are two things going on,” said Gagné. "First, you’re trying to match the letters to words that you know; and second, you’re also trying to match the individual words to a compound word that you know."

“Despite the fact this approach is not the optimal way to process a word, we still do it. Clearly, this means that we are not able to control this process,” added study co-author Thomas Spalding.

Collecting compound words

The scientists also developed a database of more than 8,000 English compound words that other researchers—whether they work in linguistics, psychology, education, or computing science and natural language processing—can use.

"Understanding how humans process compound words is very important for building robust natural language processing systems. This resource makes it easier for future studies to incorporate this element."

Gagné noted the database could also be used in health care as a way to test patients with cognitive disorders such as aphasia, and in education to help understand how children learn language.

"The more we know about how we use language, the more ways we can intervene to help or build systems that can mimic this complex process."

The study, "Detecting Spelling Errors in Compound and Pseudocompound Words," was published in the Journal of Experimental Psychology: Learning, Memory, and Cognition.

The database, "LADEC: The Large Database of English Compounds," was published in Behaviour Research Methods.