

Knowledge in Speech and Language Processing What distinguishes language processing applications from other data processing systems is their use of knowledge of language. Producing and recognizing these and other variations of individual words (e.g., recognizing that doors is plural) requires knowledge about morphology, the way words break down into component parts that carry meanings like singular versus plural. These tasks of speech recognition and speech synthesis tasks require knowledge about phonetics and phonology; how words are pronounced in terms of sequences of sounds, and how each of these sounds is realized acoustically. Sophisticated conversational agents like HAL, or machine translation systems, or robust question-answering systems, require much broader and deeper knowledge of language. Consider the Unix `wc` program, which is used to count the total number of bytes, words, and lines in a text file. Of course, `wc` is an extremely simple system with an extremely limited and impoverished knowledge of language. When used to count bytes and lines, `wc` is an ordinary data processing application. For example, HAL must know