Title
Rules and analogies in reading aloud: Hough doo peapel rede gnew wirds?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 19(0)

Author
Andrews, Sally

Publication Date
1997

Peer reviewed

Rules and analogies in reading aloud: Hough doo peapel rede gnew wirds?

Sally Andrews (S.Andrews@unsw.edu.au)
School of Psychology
University of New South Wales, Sydney 2052, Australia

Investigations of how people pronounce nonwords provide critical evidence about the nature and organization of lexical memory. Traditional symbolic models of word recognition, instantiated by the dual route model, assume that nonword pronunciations are generated from abstract rules representing the systematic relationships between letters and sounds - "grapheme-phoneme correspondence" (GPC) rules. The knowledge representations underlying generalisation performance are therefore independent of those for known lexical forms. By contrast, recent parallel distributed processing (PDP) models assume a single knowledge base acquired through experience with words. Nonword pronunciations are therefore computed by what is effectively an analogy mechanism: they are synthesised from representations for similar words.

Twenty years of empirical research on these issues has failed to conclusively distinguish between dual-route and analogy frameworks. To a large degree, this failure reflects problems in clearly operationalising what is meant by "a rule" in a manner that allows the construct to be distinguished from an analogical mechanism.

The recent development of computational implementations of both theoretical frameworks provides the precise specification necessary to resolve these ambiguities. The present research was designed to provide systematic evidence about the nature and determinants of people's nonword pronunciations and to compare this data with the outputs of Coltheart, Curtis, Atkins and Haller's (1993) Dual Route Cascade (DRC) model and Plaut, McClelland, Seidenberg and Patterson's (1996) PDP model. Both models have been demonstrated to be capable of producing plausible nonword pronunciations but they appear to differ in the specific pronunciations that they assign to particular nonwords. The PDP model tends to produce analogy pronunciations for nonwords that are similar to high frequency exception words (e.g., *mone* and *bost* are pronounced like *done* and *most*). Such pronunciations are not predicted by the rule-based procedure of the DRC model which assigns to each grapheme the pronunciation that occurs in the largest number of words, regardless of their frequency (i.e., *mone* and *bost* are pronounced like *bone* and *cost*). In the complete DRC model, pronunciations are jointly determined by cascading rule-based and lexical procedures explicit simulations of the DRC model were conducted to allow direct comparison with the human data.

In the first experiment, subjects read a long list of monosyllabic nonwords constructed to factorially manipulate the pronunciation consistency of both the initial C(C)V- and final -V(C)C units. A second experiment used a more refined nonword set and compared performance when nonwords were named alone and mixed with words.

In both experiments, at least 85% of nonwords based on inconsistent word bodies were pronounced in accordance with GPC rules by almost all of a relatively large sample of subjects. Pronunciations appeared to be primarily based on body units with minimal influence from the consistency of initial C(C)V- units. However, when nonwords were constructed from word bodies that only occur in exception words (e.g., *jourt, talf),* irregular pronunciations were at least as common as regular pronunciations. Regression analyses showed that the strongest predictor of which pronunciation people assigned was the proportion of regular word neighbours regardless of their frequency.

Thus, in contrast to the predictions of the PDP model, people almost never produce analogy-like pronunciations for words with high frequency exception neighbours. The DRC model overestimated the proportion of regular nonword pronunciations, but successfully predicted the particular nonwords that people are most likely to pronounce irregularly. The DRC model therefore provides a much better fit with people's performance than the PDP model. However, a detailed consideration of exactly <u>how</u> the DRC model achieves its successful simulation raises questions about its psychological validity.

The most fundamental of these questions concerns the basis of the model's sensitivity to the frequency of word types. The rule extraction algorithm that generates the nonlexical knowledge component of the model requires repeated presentation of the total vocabulary unweighted for frequency of occurrence in natural language. By contrast, the PDP model is trained with words weighted according to their natural language frequency. In this sense, the DRC model's ability to simulate people's apparent reliance on abstract GPC rules is a direct reflection of its training regime and completely finesses the question of how people acquire sensitivity to word types while being exposed to words weighted according to natural language frequency. In this sense, the DRC model appears to provide an accurate description of the knowledge that people use to pronounce nonwords, but brings us little closer to understanding <u>how</u> people acquire that knowledge. This goal might be best achieved by investigating how a PDP system can acquire sensitivity to word types under naturalistic training conditions.

References

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993) Models of reading aloud: Dual-route and parallel-distributed-processing approaches. <u>Psychological Review</u>, <u>100</u>, 589-608.

Plaut, D., McClelland, J., Seidenberg, M. S, & Patterson, K. (1996) Understanding normal and impaired word reading: Computational principles in quasi-regular domains. <u>Psychological Review</u>, <u>103</u>, 56-115.