Annette Kinder

The knowledge acquired during artificial grammar learning: Testing the predictions of two connectionist models

Abstract An artificial grammar learning experiment is reported which investigated whether three types of information are learned during this kind of task: information about the positions of single letters, about fragments of training strings, and about entire training strings. Results indicate that participants primarily learned information about string fragments and, to a lesser extent, information about positions of letters. Two connectionist models, an autoassociator and a simple recurrent network (SRN), were tested on their ability to account for these results. In the autoassociator simulations, similarity of test items to entire training items had a large effect, which was at variance with the experimental results. The results of the SRN simulations almost perfectly matched the experimental ones.

Introduction

Originally, artificial grammar learning (AGL) was proposed as a typical example of implicit learning: Participants in AGL experiments learn without intending to do so and seem to have little explicit knowledge about what they have learned. Nevertheless, their performance on certain tasks significantly improves with practice (for a discussion of this issue see Reber, 1989). Several experimental findings have demonstrated, however, that the matter is quite complicated: It was shown that explicit knowledge contributes to artificial grammar learning as well (Dulany, Carlson, & Dewey, 1984; Mathews et al., 1989; Perruchet & Pacteau, 1990; for an overview see Shanks & St. John, 1994).

In recent years, a further aspect of AGL that is largely independent of the debate whether learning is implicit or explicit has been brought into the focus of interest. Several attempts have been made to investigate the kind of information which is stored in artificial grammar learning (Knowlton & Squire, 1994, 1996; Perruchet & Pacteau, 1990; Shanks, Johnstone, & Staggs, 1997). Parallel to the growing interest in this issue of research, computational models of artificial grammar learning have been proposed (Dienes, 1992; Dienes, Altmann, & Gao, 1999; Servan-Schreiber & Anderson, 1990). It is crucial for the validity of these models that they make correct predictions about the kind of information which is stored during training and used for categorization. Therefore, an attempt is made in the present article to combine both lines of research: An AGL experiment is reported which investigated whether participants learned three types of information. Two computational models belonging to the class of connectionist models were tested on their ability to account for the experimental results.

In AGL experiments, letter strings are presented which have been generated using an artificial or finite state grammar. As can be seen in Fig. 1, a finite state grammar can be depicted using a number of circles (states) that are connected to each other by arrows (transitions) which are associated with letters. Letter strings are generated by following the arrows from the input state until the structure is left at an output state. Each time an arrow is chosen, the associated letter is appended to the string. Letter strings that are generated this way are called grammatical, whereas all other letter strings are called nongrammatical.

In a typical AGL experiment, participants first memorize grammatical strings. Then they are informed about the existence of a complex set of rules that determine letter order (but are not told what they are). Thereafter, they are asked to classify grammatical and nongrammatical strings. Typically, the percentage of correct responses is significantly above chance level. What do participants learn during training that enables them to correctly classify the test strings? A first answer to this question was given by Reber (1969), who...
assumed that participants acquire an abstract representation of the underlying grammar. Since then, there has been much research to challenge Reber’s view and to demonstrate that superficial features of the test strings can account for all or most of the participants’ categorization performance. The general idea is the following: Grammatical test items are in some manner more similar to the training items than are nongrammatical items. Thus, participants’ categorization behavior may rely on perceived similarity rather than on evaluating whether the items obey the rules. An exemplar-based account following this logic was proposed by Vokey and Brooks (1992), who claimed that a test string is accepted as grammatical when it is sufficiently similar to one of the learning exemplars. Actually, they demonstrated that items which were similar to one particular training item (that is, deviated from that item in only one position) were accepted with a higher probability than items which were dissimilar (that is, deviated from every learning item in at least two positions). Additionally, they found an effect of item grammaticality which was independent of the similarity effect.

A different account of AGL, which also does not refer to “rules” as constituents of participants’ knowledge, claims that the acquired knowledge is of fragments of training strings. Perruchet and Pacteau (1990), for example, argued that bigram knowledge, that is, knowledge of permissible pairs of letters, almost entirely accounts for participants’ classification performance. They tested participants’ recognition memory of bigrams which had occurred in the training strings and then used the results to simulate classification performance. In the simulations, they assumed that participants rejected any item which contained at least one unrecognized bigram. Simulated classification performance was nearly identical with the human data.

Knowlton and Squire (1994, 1996) also demonstrated that fragments (or chunks) of training strings are an important factor in AGL. They proposed a measure of chunk information (associative chunk strength) which is computed by counting how often the bigrams and trigrams of the test strings occur in the training strings. This number is divided by the total number of bigrams and trigrams the string contains. Because chunks in anchor positions are supposed to be especially important for assessing grammaticality, Knowlton and Squire (1994) suggested an additional measure which incorporates only bigrams and trigrams in first and last positions.

Knowlton and Squire (1994) demonstrated that the similarity of test items to particular training items did not affect endorsement rates when “similar” and “dissimilar” test strings were balanced for both measures of chunk strength. This finding indicates that the effect of similarity to specific training items discovered by Vokey and Brooks (1992) actually occurred due to a difference in chunk strength between similar and dissimilar test items. In accordance with this interpretation, Knowlton and Squire (1996) demonstrated that items with high chunk strength were more likely to be accepted as grammatical than those with low chunk strength.

These results can be explained by the Competitive Chunking Model, a simulation model of fragmentary learning proposed by Servan-Schreiber and Anderson (1990). This model assumes that during memorization of letter strings, a four-level chunk hierarchy is built, with the chunks containing an increasing number of letters. The bottom level, which contains representations of single letters, is thought to exist at the very beginning of the experiment. During training, the “word” level, containing representations of bigrams and trigrams, is created first. Subsequently, representations of “phrases” containing chunks of two words are created. Finally, representations of complete training strings emerge.

Representations at the same level compete with each other for strength, with stronger chunks having a higher probability of gaining even more strength. During test, strings are categorized as grammatical or nongrammatical on the basis of their “familiarity”: more familiar strings are more likely to be accepted as grammatical. The familiarity of a string depends on how strong the representations of its sub-strings are and therefore on how often these sub-strings occurred during training. In contrast to the bigram frequency account proposed by Perruchet and Pacteau and the exemplar account suggested by Vokey and Brooks, the competitive chunking model describes not only the kind of memory representations being created during AGL but also how these representations are built and strengthened.

In recent years, two types of connectionist models have been discussed as models of AGL, as well. Although these models do not incorporate the chunking principle directly, they nevertheless predict that items with higher associative chunk strength will be endorsed more readily. The models are different variants of an autoassociator and a simple recurrent network (SRN). The autoassociator is a single-layered model which consists of units, each representing a letter in a particular position (Dienes, 1992). Every unit is connected to every other one, but not to itself. During the training stage, the model gradually changes its weights in order to reproduce the input patterns presented as accurately as pos-