

Predicting the Confusions that Occur in Letter Learning

Jody S. Underwood
Department of Science and Technology Education
Technion-Israel Institute of Technology, Israel
jody@tx.technion.ac.il

1 Introduction

Marcus¹ is 27 years old, moved to Israel about a year ago, and is literate in his native language. He learned the Hebrew alphabet when he arrived. If you present an isolated letter to him, he will be able to tell you what it is. However, when he tries to identify those letters in large units, namely, words, he makes a consistent set of mistakes. He confuses:

מ (mem)	and	ט (tet)
ח (chet)	and	ת (tof)
ב (bet)	and	כ (chof)
ד (dalet)	and	ר (reish)

If you ask him why he confuses the mem and tet, for example, he will tell you that he thinks of both letters as having a flag coming out of the top, and that they look almost the same. When pressed to guess what letter one of them is, he will almost always answer, "mem".

Why should this happen? I posit that the confusions can be understood, and, through careful presentation of the letters of the new alphabet, avoided. It is possible to learn the wrong features of letters—wrong in the sense that they don't uniquely identify the characters—and correct it over time. But it takes a long time [Byrne, 1992], and in the meantime, it interferes with the other aspects of learning the language. While there are some situations in which it is better to learn from mistakes, this is not one of them. It is important not to learn them incorrectly in the first place.

Learning to read a new alphabet is one of the first problems encountered when learning a foreign language. Studies suggest that when learning to read in a foreign language, learning the skills of mapping symbols to sounds (phonological encoding) [Eckwall & Shanker, 1983; Lynch & Hudson, 1991] and letter-word recognition [Jones, 1981] are very important. But they do not directly address how to acquire those skills. Those studies that do focus on letter learning, focus on transfer and generalization of features [Byrne, 1984], whether phonological encoding helps letter learning [Koda, 1989], and where the eye moves when learning a new alphabet [Koga & Groner, 1989].

Little if any attention has been paid to the confusions that are commonly made among similar letters. Some say the acquisition of distinctive visual features—for example, the features that distinguish the letters mem and tet—is essential to the process of learning to read [Gibson & Levin, 1975; Eskey, 1987]. Again, there are no suggestions for how to achieve this goal.

My hypothesis is that when similar letters are taught together, the students are forced to see the important attributes that distinguish these letters. Otherwise, the student may focus on attributes that are not important—or worse—misleading, and it will take a long time to learn the correct features, or more precisely, to unlearn the incorrect ones.

The overall goal of this research is to find out if there is at least one sequence of a given set of symbols—for example, the letters of an alphabet—that will allow them to be learned with minimal confusion. In pursuit of this goal:

¹A real person, not his real name.

1. An exploratory study has been done in which new learners of an alphabet read texts that were a little over their heads. Briefly, they often pronounced a letter as if it were a similar counterpart.
2. A model has been developed that simulates the process by which a person learns new letters. This attempts to answer the question: Why does the order of presentation make a difference?
3. A survey has been done to determine which letters from various alphabets are considered similar.
4. An experiment has been designed that will teach letters, grouped according to the results of the survey, to show that the phenomenon occurs. I expect to find different confusions for different groupings.
5. A computer simulation of the proposed model is being developed. I expect that this simulation will be able to predict the confusions people will make based on the grouping of the input letters.

The model, the survey, and the directions that they will take us is described in the remainder of this report.

2 A model of letter learning

When a group of letters is learned together, the learner attends to the features that differentiate among the letters in the group. Let us call these original features. With practice, an automatic mapping is made from these features to the sounds associated with the letters. Without practice, they are not learned. This agrees with perceptual learning theory [LaBerge, 1975]. For example, if the letters bet (ב), mem (מ), gimel (ג), and chet (ח) are taught, the representation in memory will be something like the decision tree in Fig. 1. The tests contain the features that distinguish between the letters.

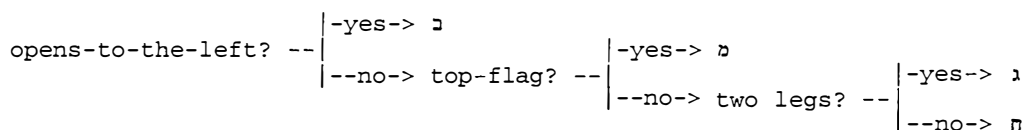


Figure 1. Features that distinguish between letters

According to perceptual learning theory, the same thing happens when a second group of letters is presented. However, this theory does not address what happens when the second group contains letters that are similar to letters in the first group. Studies in object perception allow the subject to forget about previous sets when a new one is presented. This is not realistic—learning is cumulative; it does not happen all at one time. My theory says that new features have to be integrated with the original features, and that some old mappings have to be forgotten. But it is difficult to forget that the original mappings are no longer enough to distinguish all the letters learned.

An example of how Marcus learned mem and tet will clarify this process. Marcus was presented a group of letters that contained mem, such as in Fig. 1. The characteristic that distinguished the mem from the other letters was that it had a flag on top (the original feature). Fig. 2a shows the representation of Marcus's memory, after learning this first group of letters, for deciding if a letter is a mem.

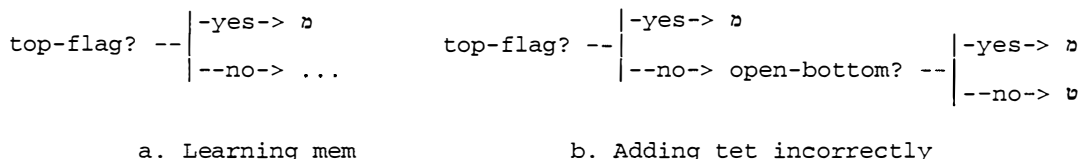


Figure 2. Decision tree when mem and tet are not learned in contrast to one another.

Marcus was then presented a second group of letters. This group contained a tet. In order to identify both the mem and tet, Marcus needed to learn an additional new feature for the mem, that is, that the mem has an opening on bottom. However, Marcus had already learned the 'top-flag' feature; to this day, whenever he sees either a mem or a tet, he thinks it is a mem. My model says that this confusion is made because there is a 'bug' in the representation of one node in Marcus's decision tree. This is shown in Fig. 2b.

Each node in these decision trees contains two items: a decision and a test. Valid states for a node are:

```

decision = NIL      or      decision = beta
test = alpha        test = NIL

```

However, when a node is changed incorrectly from the latter to the former, it can end up in a third, illegal state:

```

decision = beta
test = alpha

```

When identifying a letter without time pressure, the test leads to the next node in the tree, and a correct decision will be made. But under pressure, the current node's decision will be accepted as the letter.

The node that contained the decision mem in Fig. 2a needed to have two things done to it: add a test (open-bottom), and remove the decision (mem). However, the decision was not removed because Marcus previously learned that 'top-flag' was the distinguishing feature for a mem. The node mistakenly continued to make a decision, namely mem, when seeing a letter with the feature 'top-flag'. My preliminary studies show that it is hard to unlearn this; the solution lies in ordering/grouping so as to avoid the problem in the first place.

I hypothesize that the confusion explained in the model occurs most often when reading words. When reading words, you want to quickly decide what all the constituent letters are. When identifying letters individually, there is less pressure to quickly make a decision because other letters are not demanding your attention. It is possible that the buggy decision node contains an item that alerts the person that a 'test' also exists, but it only appears when there is no time pressure for identifying the letter. This observation needs to be more systematically tested. If it is true, then a modification to the model will need to be made.

I hypothesize that the first letter learned is always the decision reached when a confusion of this type occurs. If the confusion is more symmetric, that is, either of the confused letters is chosen randomly, then the model needs to be changed to reflect it. This will be tested in an experiment, which is described later.

From the model we can infer that (1) novices will vary in the way they evaluate the similarity of a pair of letters (they have not yet decided which features are important to focus on); and (2) experienced users of an alphabet will have different judgements of similarity for any given pair of letters (they focus on different features of the letters, based on how they learned them). An informal study showed that these phenomena actually occur.

3 Letter similarity survey

A survey was run to find out what letters people found similar for the purpose of grouping the letters in the experiment. People were asked to rate the similarity of pairs of letters on a scale of 0 to 5 (see Fig. 3). Two character sets were evaluated—the ASL fingerspelling alphabet (hereafter referred to as ASL), and the Japanese Katakana syllabary. The survey was run over the World Wide Web and required a graphical browser like Netscape. For a description of how the survey was run, see [Underwood, in press]. The participants included people who read newsgroups about: teaching English as a foreign language, sign language, Japanese linguistics, foreign language laboratories, and linguistics.

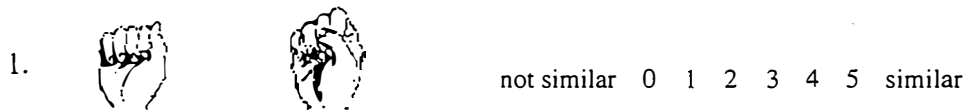


Figure 3. Survey question: rating of the ASL letters "a" and "s", respectively.

There are 325 pairs of letters in the 26-letter ASL alphabet. These pairs were distributed into 5 files of 65 pairs each to shorten the task for each evaluator. People were generally not willing to do the survey if it would take more than 10 minutes. One of the files was randomly chosen and sent to each evaluator. Because of this, some files were rated more often than others. Each pair in the ASL alphabet was rated on average by 16 people (ranging from 7 to 21). Three types of people rated the pairs—on average, 8 who have Never seen them, 4 who have Played-with them, and 4.5 who are Serious users of the alphabet. Not enough people have rated the Katakana syllabary yet to come up with coherent results.

People in the Never and Serious groups rated the same pairs of letters very differently. This hints at two things: (1) different features are attended to by each group, and (2) people who have seriously used ASL have come to think of the letters differently than they did in the beginning. The graph in Fig. 4 shows pairs of letters that the Serious group did not find similar and that the Never group did find similar.

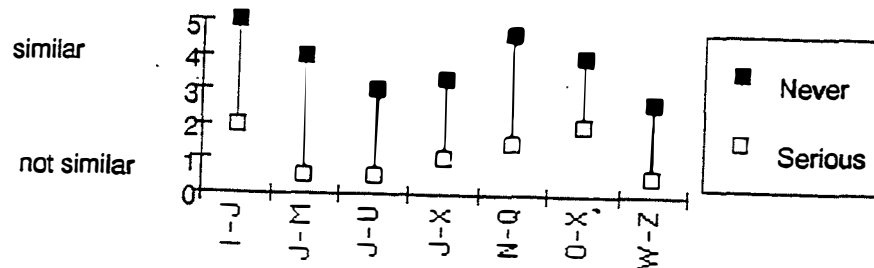


Figure 4. Divergence of Never and Serious ratings for certain pairs of ASL letters

In addition, people in the Serious group rated the same pairs of letters differently more often (33 pairs with std dev ≥ 2) than people in the Never group (5 pairs with std dev ≥ 2). This implies that Serious individuals focus on different features in the letters than Never individuals.

4 Future Directions

Experiment: Does the grouping of letters affect learning?

An experiment will be run to see if the grouping of similar letters affects learning. The subjects will be adults who are literate in their native languages and have no familiarity with the new alphabet. They will be trained on the letters by a computer program. The experiment will be run on two different alphabets.

There will be two experimental groups, as shown in Tab. 1. Both groups will learn two sets of letters. The order of the sets will be different for half the subjects in each condition. "Similar letters" means that each set of letters contains letters similar to each other. "Dissimilar letters" means that each set of letters contains letters not similar to each other, with similar counterparts in the other set. In the end, both experimental groups will have learned the same letters. The groupings will be based on the results from the letter similarity survey. Sample sets of letters using the Roman alphabet are shown in Tab. 1.

1. Similar letters	set1: DOPRMNUH set2: TISZVYFE
2. Dissimilar letters	set1: DPMUTSVF set2: ORNHIZYE

Table 1. Experimental groups

The letters will be taught over the course of two days. After the first session, the subjects will explain how they identified each of the letters. These explanations will point to features that are attended to for all the letters. The subjects will be tested after learning the two sets. The subject will read sentences made up of word-like units, made up of letters from both sets, that are displayed on the screen in sentence format (a series of words)—for example: DUM TOM IS SORN. Words and non-words in sentences, rather than individual letters, will be used for testing in order to create a time pressure to identify the letters. This puts the subject in a more natural reading situation, which is more likely to produce the confusions found in everyday identification of letters on which this research is focusing.

The model predicts that the first experimental group will learn the letters best (fewest mistakes in testing). The second group will have problems because they will not be induced to focus on those features that would distinguish the letters from their similar counterparts.

Predicting letter learning confusions

Decision trees are used to discriminate between objects by asking questions. Simon and Feigenbaum [1964, 1984] used a decision tree, EPAM, to model object discrimination in long term memory. They claim that discrimination is done by memory structures that are like decision trees. Decision trees are used in areas such as artifact dating, pattern recognition, and speech recognition [Ur, 1994].

I will use decision trees to simulate the model that is described earlier. The feature descriptions of the letters of an alphabet will run through a program, grouped in different ways, as was done in the experiment. The output will be a decision tree. Each decision tree will be structured in such a way that it will be possible to predict exactly the confusions that will be made by subjects in the experimental groups. This means that in some cases there will be "buggy" internal nodes, and in some cases there will not. A correct decision tree will be able to generate the order and groups to teach letters that will create a minimal amount of confusion. It can also manipulate the order so that particular confusions will be made.

Letter descriptions

The letters of the alphabets will be described in terms of the features that distinguish every pair of letters. For example, a difference between the ASL letters "a" and "s" (see Fig. 3) is that "a" has a thumb extended upright to side of hand and "s" has a thumb bent across hand. In fact, both letters also have all four fingers bent over against the palm. But this feature is not important for distinguishing between these two letters. If all letters in the alphabet were distinguished only by the position of the thumb, then that is the only feature that would be necessary to attend to in order to learn the entire alphabet. A learner of the ASL alphabet would not be aware that this was not the case until more letters were seen.

These descriptions will be produced by humans, based on data taken from the experiment. There has to be a way to simplify this arduous task, otherwise this decision tree software will not be useful.

5 Theoretical background

There are a few theories that explain how to view letter learning as feature-based recognition. These include prototype theory, skill acquisition, analogy and similarity, and object perception. Only the latter will be addressed here; a preliminary review of the other theories has been done, but it is not ready to be presented.

Object perception

Research in the representation of object perception shows that the features of letters that are perceived are not only at the lowest level (structuralist) or at the highest level (Gestalt), but also lay somewhere in between, and most likely are organized hierarchically [LaBerge, 1975; Palmer, 1977]. LaBerge [1975] says that perception of objects occurs in stages:

1. Feature discovery stage. Attention is paid to features that distinguish one object from another—for example, if a letter has an opening. Results from Palmer's [1977] experiments agree.
2. Coding. Features are combined with the relations between them—for example, 'flag' is a feature, 'on top' is a relation.
3. Automatic coding. Features and relations are chunked into single units.

My model of letter learning generally agrees with these guidelines. My model extends this theory by focusing on the confusions that are made during the iteration of the stages when more features need to be attended to in order to discriminate between letters.

6 Modelling and Simulation: Related work

There have been quite a few attempts at *modelling letter learning*. Some researchers use hierarchical clustering of features. Palmer [1977] built a model that successfully predicts the features that people consider salient in the domain of artificial stick-like orthography. It does not address more general objects—for example, letters with curves—and has a limited definition of "feature" (a 3-segment portion of a stick figure). Palmer's work only addresses salient features when one "letter" is seen, and not when more than one item is seen at one time. The current research addresses this, and in addition, it uses real letters in all their complexity.

Many researchers have used neural networks to *model letter recognition* [McGraw et al., 1994; Fukushima, 1992; Seidenberg and McClelland, 1989]. Neural networks are good at generalizing from noisy data, that is, discovering prototypes. But they get confused when two letters are similar—they think one is a noisy version of the other. Using decision trees for the simulation in the current research will help to deal with this problem.

References

- Eckwall, E.E., & Shanker, J. (1983). *Diagnosis and remediation of the disabled reader*. MA: Allyn and Bacon.
- Lynch, B.K. & Hudson, T. (1991). EST reading. In M. Celce-Murcia (Ed.) *Teaching English as a Second or Foreign Language*, 2nd Ed., 216-232. Heinle & Heinle.
- Byrne, B. (1984). On teaching articulatory phonetics via an orthography, *Mem & Cog*, 12 (2), 181-189.
- Byrne, B. (1992). Studies in the acquisition procedure for reading: Rationale, hypotheses, and data. In P. Gough, L. Ehri, R. Treiman (Eds.) *Reading acquisition*. Hillsdale, NJ: LEA, Ch.1, 1-34.
- Eskey, D. (1987). Conclusion. In J. Devine, P.L. Carrell, & D. Eskey (Eds.), *Research in reading English as a second language*, 189-192, Washington, D.C.: TESOL.
- Fukushima, K. (1992). Character recognition with neural networks. *Neurocomputing*, 4, 221-233.
- Gibson, E.J., & Levin, H. (1975). *The psychology of reading*. Cambridge, MA: MIT Press.
- Jones, E.V. (1981). *Reading instruction for the adult illiterate*. Chicago: American Library Association.
- Koda, K. (1989). Effects of L1 orthographic representation on L2 phonological coding strategies, *Journal of Psycholinguistic Research*, 18 (2).
- Koga, K., and Groner, R. (1989). Intercultural experiments as a research tool in the study of cognitive skill acquisition: Japanese character recognition and eye movements in non-Japanese subjects. In H. Mandl and J.R. Levin (Eds.), *Knowledge acquisition from text and pictures*. North-Holland, 279-291.
- LaBerge, D. (1975). Perceptual learning and attention. Chapter 5 in W.K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 4), 237-273. Hillsdale, N.J.: Erlbaum.
- McGraw, G., Rehling, J., & Goldstone, R. (1994). Roles in letter perception: Human data and computer models. CRCC-TR 90, Department of Psychology, Indiana University, Bloomington, Indiana.
- Palmer, S. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9, 441-474.
- Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming, *Psychological Review*, 96 (4), 523-568.
- Simon, H.A. & Feigenbaum, E. (1964). An information-processing theory of some effects of similarity, familiarization, and meaningfulness in verbal learning. *J. Verbal Learning and Verbal Behavior*, 3, 385-396.
- Simon, H.A. & Feigenbaum, E. (1984). EPAM-like models of recognition and learning. *Cog Sci*, 8, 305-336.
- Underwood, J.S. (in press). Data Collection on the Web, *CAELL Journal*, 1996.
- Ur, S. (1994). Decision trees, orders, and order restrictions. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.