_lognostics
tools for vocabulary research

Matrix models of vocabulary acquisition
Paul Meara ***Swansea University***

The traditional way of comparing the effectiveness of different ways of teaching vocabulary in L2 is essentially very straightforward. In its simplest form, we take two groups of learners, carefully controlled for obvious variables likely to effect the outcome. We then expose one group to learning method A, the other group to learning method B, and compare the number of words learned by group A with the number of words learned by the B group. The appropriate statistical texts then allow us to decide whether the differences we find are significant. In practice, of course, things are never this simple. There is, for instance, the much vexed question of how we define "knowing a word". There is also the problem that it may be difficult, or impossible, to find a testing procedure which is truly neutral between methods. A single test format that could evaluate the efficiency of the keyword method and learning words from context, for instance, is hard to imagine.

Nevertheless, despite these major difficulties, there are circumstances where people do want to compare the efficacy of different methods, and a number of experimental research papers which claim to do just this have appeared in recent years. Occasionally, these papers do rather more than what I have outlined above, and report not just the initial differences between the groups, but also the way these differences persist over time, or diminish over time as the case may be. Such changes are obviously important: a huge initial difference that disappears after 24 hours is not worth making a big song and dance about; on the other hand, a very small initial difference which increases as time goes by, needs to be taken more seriously.

Surprisingly, perhaps, not much attention has been paid to the longer term benefits which can be attributed to the various learning methods. The only studies I know of which address this question are laboratory studies, where "long-term retention" is studied over ten days or so, rather than the several months or years that we are really interested in as teachers or language learners.

In this paper, I want to propose a rather more sophisticated approach to this question. Before we can do this, however, we need to look at a simple mathematical model of the process we are trying to describe. Let us imagine that we have a situation where we teach an individual learner (let us call him Pedro) 100 L2 words, and we are interested to know how many words he will retain after six months. No learner is perfect, so let us imagine that at the time of the first testing period, T1, Pedro has managed to remember 90 of these words, but has failed to remember the other ten. We know that over the next few days, he is going to forget some of words he knows now; he might also remember some of the words that he has already forgotten. Let us pretend for the sake of argument, that we actually know how likely it is that Pedro will forget one of the 90 words he knows, or spontaneously regenerate a forgotten word. Let us say, that over seven days, there is a 1 in 10 chance that he will forget a word, and a 1 in 10 chance that he will spontaneously recover a forgotten word. We can map out this set of probabilities as shown in Table 1.

Table 1 shows that Pedro starts off knowing 90 percent of the words we have taught him. Between T1 and T2 there is a 90 percent chance of him retaining each of these words, and a 10 percent chance of him forgetting each word. There is also a 10 percent chance that a forgotten word will regenerate, and a 90 percent chance that a forgotten word will stay forgotten.

Table 1: probability of transition between 2 states

transition matrix	**k2**	**f2**
k1	.9	.1
f1	.1	.9

	K	:	**F**
T1	90	:	10

k1=words known at T1 f1=words forgotten at T1
k2=words known at T2 f2=words forgotten at T2

Given these data, we can calculate what state Pedro's vocabulary will be in at the second test time, T2. He should know:

(90 * .9) + (10 * .1) = 82 words

and he should have forgotten the rest of the words:

(90 *.1) + (10 * .9) = 18 words.

This gives us:

	K	:	**F**
T2	**82**	**:**	**18**

We can use the same method to predict what should be the state of Pedro's vocabulary at test time 3, by taking the new start state (82 known words, 18 forgotten words) and multiplying these figures by the transition matrix. And, in fact, we can predict the state of

Pedro's vocabulary knowledge at a number of subsequent testing times, as shown in the left hand column of table 2.

Table 2 The effect of iteration a transitional probability matrix

Transition Matrix

	k2	**f2**
k1	.9	.1
f1	.1	.9

test time	**K : F**	**K : F**	**K : F**
T1	90 : 10	80 : 20	30 : 70
T2	82 : 18	74 : 26	34 : 66
T3	75 : 24	69 : 30	37 : 62
T4	70 : 29	65 : 34	39 : 60
T5	66 : 33	62 : 37	41 : 58
T6	63 : 36	59 : 40	43 : 56
T7	60 : 39	57 : 42	44 : 55
T8	58 : 41	46 : 43	45 : 54
T9	56 : 43	55 : 44	46 : 53
T10	55 : 44	54 : 45	47 : 52
T11	54 : 45	53 : 46	47 : 52
T12	54 : 45	52 : 47	48 : 51
T13	53 : 46	52 : 47	48 : 51
T14	52 : 47	51 : 48	49 : 50
T15	51 : 48	51 : 48	49 : 50
T16	50 : 48	51 : 48	49 : 50
T17	50 : 49	50 : 49	49 : 50
T18	50 : 49	50 : 49	49 : 50
T19	50 : 49	50 : 49	49 : 50
T20	50 : 49	50 : 49	50 : 50

(the totals fail to add up to 100 due to rounding errors.)

This table reveals a rather surprising phenomenon. It looks, at first, as if Pedro's retained vocabulary is going to disappear completely in time, but this isn't what happens. Instead, the system reaches an equilibrium point after about 15 iterations, reaching a state where the number of items forgotten is balanced by the number of items that spontaneously regenerate. In this example, Pedro retains about half the words we originally taught him.

What factors determine this equilibrium level? The middle column of table 2 shows what happens if we start from a different initial state, with Pedro retaining only 80 percent of the original 100 words at test time T1. Surprisingly, the system settles down into the same equilibrium level as our first example. In this second example, Pedro also ends up remembering about 50 percent of the original 100 word set. The right hand column of table 2 shows an even more dramatic example, where Pedro remembers only 30 percent of the original 100 words at test time T1 . Amazingly, in this case, spontaneous regeneration causes the number of words he knows **increase**, and again after 20 or so iterations the system eventually settles down into an equilibrium which is the same as the equilibrium that we found with the two previous examples.

In Table 2, we varied the number of words that Pedro remembers at test time T1. Table 3 shows what happens if we vary not the number of words known at T1, but the matrix of transitions between the states. Each of the examples assumes that Pedro knows 90 of the words at test time T1, but makes different assumptions about the probability of remembering or forgetting these words. Here again, we see that the systems eventually settle down into an equilibrium after 15 or so iterations, but this time, the equilibrium points are different.

The matrix at the head of the left most column in table 3 resembles the matrix we examined in table 2, in that the probability of spontaneous regeneration remains at one chance in 10. The likelihood of a known word being retained is lower than in our previous examples, however. Not surprisingly, this produces a lower equilibrium point than our original matrix. In the second matrix in table 3, we have increased probability of spontaneous regeneration relative to our original matrix, and we find that a gratifyingly high equilibrium point emerges. The third matrix in table 3 is perhaps more like what happens in real life. This matrix allows for a modest retention rate from one week to the next, coupled with a very small chance of spontaneous regeneration. The long-term equilibrium point of this system is only 11 percent.

What we have shown here is that finite state transition matrices of this kind have an interesting mathematical property: their equilibrium points are completely independent of the starting state of the system. It is the transition matrix, not the initial starting point, that determines the equilibrium level. The starting point affects how long it takes for the system to reach its equilibrium level, but in the long-term, it is not relevant to the final state of the system.

We began this paper by asking: how can we assess the effectiveness of two different vocabulary acquisition programs? The mathematical arguments we have just reviewed

Table 3
How the equilibrium point is determined by the transition matrix

Transition Matrix		**k2**	**f2**	**k2**	**f2**	**k2**	**f2**
	k1	.8	.2	.9	.1	.6	.4
	f1	.1	.9	.3	.7	.05	.95

test time	**K : F**	**K : F**	**K : F**
T1	90 : 10	90 : 10	90 : 10
T2	73 : 27	84 : 16	54 : 45
T3	61 : 38	80 : 19	34 : 65
T4	52 : 47	78 : 21	24 : 75
T5	46 : 53	76 : 23	18 : 81
T6	42 : 57	76 : 23	15 : 84
T7	40 : 59	75 : 24	13 : 86
T8	38 : 61	75 : 24	12 : 87
T9	36 : 63	75 : 24	11 : 88
T10	35 : 64	75 : 24	11 : 88
T11	34 : 65	75 : 24	11 : 88
T12	34 : 65	75 : 24	11 : 88
T13	34 : 65	75 : 24	11 : 88
T14	33 : 66	75 : 24	11 : 88
T15	33 : 66	75 : 24	11 : 88
T16	33 : 66	75 : 24	11 : 88
T17	33 : 66	75 : 24	11 : 88
T18	33 : 66	75 : 24	11 : 88
T19	33 : 66	75 : 24	11 : 88
T20	33 : 66	75 : 24	11 : 88

(the totals fail to add up to 100 due to rounding errors.)

suggest that the traditional method -- a one off post-test administered immediately after learning -- completely misses the real point at issue. Even when an immediate post-test is combined with a later retest, the resulting data is still inadequate for our purposes. The crucial data is not the number of words learners know at an arbitrary point in time, but the transitional probability matrix that defines the eventual equilibrium level. It looks as though the answer to our question lies in investigating the structure of the transitional probability matrices induced by our teaching programs rather than the more obvious raw data that they generate.

Before we can take this interesting idea any further, however, we have to face a number of serious objections to the kind of approach I have outlined here. We will deal with the theoretical objections first.

The basic theoretical objection is that the simple two state matrix model proposed here is fundamentally inadequate to describe what happens when we acquire vocabularies. It is inadequate in two main ways. Firstly the simple distinction between words you know and words you don't know grossly misrepresents the real world. We really need to distinguish between: a) words you know really well; b) words you know partially; c) words you know you knew once, but can't remember anymore; d) words you have totally forgotten but would recognize if your memory was jogged; and e) words that you never really learned at all. This objection is easily met. There is no a priori reason why we should not look at transitional probability matrices involving three or more states. The mathematics of these models is more complicated and there may be serious practical difficulties in deciding which words are deemed to fall into each category -- two good reasons for working initially at least with simpler models. In principle, however, the basic argument is the same: if you have a multistate system, coupled with a probability matrix showing the chances of an item moving from one state to another over a given period of time, then we would expect the system to reach an eventual equilibrium state in which the proportion of items in any one of the states remains constant.

The second theoretical argument is more difficult to deal with. The position I have advanced here, clearly depends on the assumption that the values in the transitional probability matrix remain constant, and do not change from between test events. If the values change systematically, then we could predict the long-term outcome of the system, but the calculations would be a great deal more difficult and complex.. On the other hand, if the values change in an unpredictable way, then clearly, we cannot predict the long-term outcome of the system. I don't have any real answer to this objection. It seems to me highly likely that the transitional probability matrix is unstable, and changes with an individual learners personal circumstances. For example, it might be the case that the probability of retaining newly learned words increases as a learner's vocabulary increases in size. In the last analysis, however, this objection is an empirical question that needs to be investigated. For the moment, we can probably work on the simplifying assumption that the values in a subject's transitional probability matrix are fairly stable, and do not change very much over short periods of time.

This then brings us back to the practical objections against using these transitional probability matrix models. There are three main objections to the type of approach I have outlined here. The first objection is that there is no obvious way of determining what any individual student's transitional probability matrix looks like. For instance, in my discussion, I have not specified how much time passes between one test event and the next. It might have been a day, a week, a month, or whatever. Clearly, it makes a difference which unit of time we choose: forgetting rates over 24 hours are not the same as forgetting rates over 24 days or 24 months. My hunch is that a week is a good time period to use. Most educational institutions work on a seven-day cycle, which means that it is relatively easy to organize test-retest sessions at this distance apart. Forgetting over a week is substantial, but far from complete. My suggestion, therefore, is that it would be sensible for people working in this field to adopt a convention than the basic unit of time is seven days. It may turn out that this time unit is inappropriate, but adoption of the convention as a convenient simplification would at least mean that work from different sources could be evaluated within a common framework.

The second objection is more substantial. Assuming we agree that the interesting data is the underlying transitional probability matrix, rather than the raw scores, how do we determine what figures should be entered into the matrix? There are number of possible answers to this objection, but again, it seems that the simplest practical solution is to adopt a reasonable convention. The most obvious solution would be to test Subjects immediately after the initial learning period, and again one week later: we can then calculate the proportion of words the Subjects know at the end of the initial learning, how many of them are retained T2, how many forgotten at T2, and so on, and enter these figures in the appropriate cells of the matrix. In practice, however, we have found that this method is not a good predictor of the final equilibrium state: it tends to overestimate the rate of forgetting, and to underestimate spontaneous regeneration. Our experience suggests, that the transition patterns from T2-T3, T3-T4, T4-T5 and so on *are* fairly stable and any one of these is a better predictor than the T1-T2 transition. The obvious convention to adopt, then, is to use the transitional probabilities from T2-T3 as the basic matrix. Note, however, that this implies a three stage testing program, which might be awkward to implement in practice. It might, be possible to adopt a difference convention: that we might use the T1-T2 matrix with appropriate compensating adjustments. At the moment, however, I do not have enough data to be in a position to suggest what these appropriate adjustments might be.

The third practical objection to the programme I am suggesting here is the question of objective, neutral tests of vocabulary knowledge. Again, there is no obvious answer to this difficulty. We have found, however, that learners are actually fairly conservative in their judgments about whether they know words or not, and they tend to underestimate their abilities rather than to overestimate them. It is possible to exploit this tendency by giving subjects a list of test words intermingled with a list of nonexistent words which closely resemble the real test words. The subjects are then simply asked to mark all the words they know. Subjects rarely mark the non-words in this situation, and this suggests that they are being basically honest about their knowledge of real words. In cases when more than a few percent of the non-words are marked, we can infer that the subjects are overestimating their knowledge of real words, and appropriate adjustments to their real word scores can be made.

Table 4
Example of a YES/NO vocabulary test

Look through the French words listed below. Cross out any words that you do not know well enough to say what they mean.

vivant	trouver	magir	romptant	laibure
mélange	livrer	ivre	fombé	sid
moup	vion	lague	inondation	roman
soutenir	siècle	torveau	prêtre	chic
repos	ganal	harton	toule	ornir
goûter	foulard	exiger	avare	cérise
étoulage	écarter	mignette	jambonnant	papiment
démenager	poignée	équipe	missonneur	confiture
ajurer	barron	clage	toutefois	gôter
leusse	cruyer	hésiter	surprendre	ponte

An example of a test in this format will be found in table 4. There are a number of problems with tests of this type -- notably that they measure passive to vocabulary, rather than active vocabulary skills. However, they do have some important practical advantages -- ease of construction, simplicity of assessment, time necessary for completion, possibility of using large sampling rates, and so on -- which seem to outweigh most of the theoretical disadvantages. On balance, this type of test looks like the nearest thing we have to a practical minimal test of vocabulary knowledge in an L2, and its widespread adoption as a research tool would be worth consideration. Further discussion of this type of test instrument will be found in Meara and Buxton (1987) and Read (1988).

Conclusions

In this paper, I have suggested that traditional ways of assessing the effectiveness of vocabulary testing techniques suffer from a major theoretical flaw: they concentrate on superficial phenomena, and neglect the underlying structure of these phenomena. I have suggested that the underlying structure can be seen as a transitional probability matrix, and that the characteristics of these matrices in real L2 learners are worth some detailed study. However, it is unlikely that such study will get very far unless we adopt some common standards for research. I have suggested three common standards which might help to make diverse research programs more compatible. These are:

1: adoption of a week as the basic unit of time;

2: adoption of the T2-T3 transition matrix as the standard datum;

3: adoption of a neutral, minimal vocabulary assessment method.

I must admit that I feel a certain unease in making a suggestion of this sort. Research thrives on anarchy, and tends to become routine and boring where external controls are imposed. However, my experience in editing the two bibliographical source volumes published by the CILT (Meara 1982, 1987) has convinced me that the field as a whole would be more coherent and easier to interpret if a few, commonly agreed standard conventions were adopted. I hope that colleagues who (quite rightly) object to authoritarianism in research will be able to accept these suggestions in the spirit in which there are intended.

References

Meara, PM
Vocabulary in a second language. London: Centre for Information on Language Teaching. 1982.

Meara, PM
Vocabulary in a second language, volume 2. London: Centre for Information on Language Teaching. 1987.

Meara, PM and B Buxton
An alternative to multiple-choice vocabulary tests. *Language Testing* 4,2(1987), 142-154.

Read, J
Measuring the vocabulary knowledge of second language learners. *RELC Journal* 19,3(1988) 12-25.

Notes

An earlier version of this paper first appeared in *AILA Review* 6(1989), 66-74.