It has been claimed that error analysis (EA) has two broad aims and two levels of application: pedagogical (relevant to syllabus design and second language teaching) and psycholinguistic (relevant to language learning studies). At the moment, EA's pedagogical claims are stronger than its psycholinguistic ones. In its early days, EA defined its object of study as the native and the target language and the points of similarity between them. This was based on the notions of contrastive analysis and interference. From criticism of this approach grew the notion of interlanguage. This, too, is open to criticism. It is proposed that a new approach is needed toward the learner's knowledge and use of the target language, one which is neither purely psycholinguistic nor pedagogical, but which incorporates a sociolinguistic element. The suggested approach describes the learner's language in terms of his or her "linguistic repertoire," which includes the native language and the target language. (AM)
It is customary to claim that Error Analysis (EA) has two broad aims, and two levels of application: pedagogical and psycholinguistic. Its pedagogical relevance is to the design of syllabuses and teaching materials, particularly for remedial courses; and it also provides feedback about the efficacy of existing teaching methods and materials. This pedagogical application of EA thus amounts to an elaborated and extended form of diagnostic testing.

The second application is psycholinguistic: particularly via longitudinal studies, it may throw light on the way languages are learned in general, and on the relationship between the acquisition of a native language and the learning of a foreign language. The psycholinguistic feedback it may provide could in turn affect the linguistic description of the language(s) concerned. This side of EA has been called performance analysis (Svartvik 1973).

Although diagnostic testing is nothing new, EA as a separate branch of applied linguistics is scarcely a decade old. It is thus not really surprising that its pedagogical claims stand, at the moment, on much firmer ground than its psycholinguistic ones. Problems arise in particular from current definitions of precisely what EA should study, and from attempts to find explanations for errors. In addition, the links between theories of language learning and the general psychology of learning are still somewhat tenuous.

After discussing some of the theoretical problems that have arisen from within EA itself, I shall try to show that some loose ends can be drawn together if a sociolinguistic dimension is introduced into the EA model.

THE OBJECT OF STUDY OF ERROR ANALYSIS

The first (applied linguistics) approach to EA was that based on the notion of interference, the psychology of negative transfer, and Contrastive Analysis (CA), stemming primarily from the work of Lado (1957).
The object of study was defined in linguistic terms: the native language (LI) and the target language (TL) of the learner, and particularly the points of similarity and difference between them.

There has been a great deal of discussion and criticism of this view, in its extreme forms at least, and I need not repeat it here. (See e.g. Nemser 1971, Dulay and Burt 1974, James 1971.) The consensus view nowadays seems to be that LI interference can certainly explain some errors—perhaps about one third for adult learners (George 1972) — but there are many errors that cannot be attributed to LI interference, and other explanations are needed for these.

Out of the reaction against the CA approach there has grown the hypothesis-forming theory, also known as the interlanguage hypothesis (Selinker 1972, Frith 1975), the LI-learning vs L2-learning hypothesis (Dulay and Burt 1974), or simply the cognitive approach. There has so far been far less criticism of this approach, and I shall therefore discuss some of the problems it raises in rather more detail.

First proposed by Corder (1967, 1971), it has since been developed by Richards (1971), Selinker (1972), Nemser (1974), and many others.

This new view of EA brought about a shift of emphasis from formal causes to efficient causes (Kellerman 1977), from the languages to the learner himself: the object of study is defined now more in psycholinguistic than linguistic terms.

According to Corder, EA should study the transitional competence of the learner, who is said to speak an idiosyncratic dialect comparable to the languages of infants, poets and aphasics (1971). He argues that EA should be concerned with the whole of the learner's use of the TL, not just the errors, because "every sentence is to be regarded as idiosyncratic until shown to be otherwise" (Corder 1971:155).

These terms, and, more importantly, the concepts underlying them are not entirely satisfactory. Firstly, the term 'transitional' seems to suggest a movement, presumably from LI to TL—an unreasonable view, since the learner does not leave his LI behind. Further, this movement is understood to be a succession (presumably infinite) of intermediate stages (Nemser's approximate systems). Yet any language changes, any competence is transitional, not just a learner's. Maybe a learner's language changes itself, is more transitory, but this is a relative, not an absolute difference. What EA needs is a concept which captures this relative differ-
There are also problems with the term 'competence'. Zydatis (1974) has argued convincingly that although such a concept may be useful from the psycholinguistic point of view, it is irrelevant to the pedagogical view of EA: teachers must be able to talk about errors, about unwanted forms; but these are not relevant terms in a description of competence, which, by definition, is at any one time complete and perfect in itself.

The concept of an idiosyncratic dialect is also open to criticism on similar grounds. Seeing the learner's use and knowledge of the TL as a 'dialect' will work for psycholinguistic EA, just as 'competence' will, but it is scarcely an appropriate term for the pedagogical approach: if the learner's language is a dialect in its own right, we are in no position to talk about errors in this dialect. 'Idiosyncratic', on the other hand, makes sense at the pedagogical level: certain forms can usefully be idiosyncratic with respect to the TL, but the term is irrelevant for the psycholinguistically oriented approach, which studies the learner's language as an independent system, in itself no more idiosyncratic than any other language system.

Selinker's (1972) term for what EA should study is interlanguage (IL). This has become perhaps the most common label for the learner's knowledge and use of the TL, but in some respects this is also misleading. It is often pictured as a circle, overlapping on each side the circles of the IL and the TL, which suggests that the IL system is made up of elements drawn partly from the LI system, partly from the TL system, and partly from neither. Yet the IL itself, like a 'dialect' or a 'competence' is defined as a "separate linguistic system", this is not identical to the LI, of course, but nor is it identical to the TL. This is because a given set of utterances in an IL are not identical to the "corresponding set of utterances which would have been produced by a native speaker of the TL had he attempted to express the same meaning as the learner" (Selinker 1972; 1974:35).

This view of the IL as a separate system implies that the IL-speaking learner never actually speaks the TL, which is sociolinguistically rather strange. When I, as a native speaker of English, speak Finnish, there may be errors in my Finnish, and some of these errors may be partly due to interference from English; nevertheless I am, I claim, speaking Finnish, albeit in a reduced and somewhat irregular form. After all, a
Fi will understand me, more or less, and a monolingual Englishman will not. The difference between my Finnish and a Finn’s Finnish is, rather, a relative one.

A further problem with the concept of the IL as a separate system is: precisely what is meant by ‘system’? Corder (1971) and Nemser (1971) also stress the systematic nature of the learner’s use of the TL, but Frith (1975) has pointed out, the concept of systematicity, as used in EA, has yet to be adequately defined. It must presumably be a relative concept, for a start. If by ‘systematic’ is meant ‘very regular, and so predictable’, then it must be admitted that many errors - even those of advanced learners' - appear to be random, not predictable. Moreover, although it is often possible to predict areas in which errors are more likely to occur, it is much more difficult to predict exactly what form an error will take.

However, even granted that much of the IL is fairly regular, relatively little of this regularity consists of systematic errors; most of it, for learners past an absolute beginner’s stage at least, consists of the regularity of the TL system itself. For this reason too, it seems unnecessary to postulate a separate IL system.

I suggest, therefore, that we need a different angle of approach to the learner’s knowledge and use of the TL: one which is neutral between the psycholinguistic and pedagogical views; which would give more weight to the flexibility and growth of the learner’s command of the TL; and which would allow us to say that the learner does in fact speak the TL.

THE PROBLEM OF EXPLANATION

In terms of learning theories, the approach adopted by Corder et al. rests on a cognitive view of language learning; they argue that learning is not the habit formation of behaviourist theory (and early Contrastive Analysis), but rule-governed behaviour. Much emphasis is given to this distinction.

However, no satisfactory explanation has yet been offered of what the difference between habits and internalized rules really is. Carroll (1968: 114) doubts whether there is one at all:

I am not convinced... that there is any real difference between a 'habit' and a 'rule', or between a 'response' and a 'rule-governed performance'.
Although this distinction seems to me to be a false one, the stress given to the learner's active participation in the language learning process is valuable, and any theory of language learning must certainly take into account the various cognitive strategies of the learner. Only in this way can errors be explained.

LI interference is recognized to be one source of explanation: Kellerman (1977) presents a fairly detailed analysis of LI transfer, treating it as a cognitive strategy in its own right. However, classifications of other strategies or explanations that have so far been proposed are woolly in the extreme. Richards (1971), for example, distinguishes "overgeneralization" from "ignorance of rule restrictions" and "incomplete application of rules". Quite how these are to be distinguished I fail to see: the overgeneralization of one rule is surely the result of ignorance of the restriction of another rule, and shows that a further rule has been incompletely applied.

Zydatiss (1976:338) points out that current classifications of explanations also equate like with unlike.

On the one hand we have psychological mechanisms like LI-interference and overgeneralization (which in fact are inferences from observable data), on the other hand we have a linguistically oriented product analysis (which would incorporate categories like 'ignorance of rule restrictions' and 'incomplete application of rules'). We should also distinguish between sources of error, examples of which are certain teaching procedures like overteaching or incorrect presentation in the textbook, and learning strategies, to which we may want to assign hyper-correction, simplification, and Levenston's (1971) notions of over-indulgence and under-representation.

Selinker (1972) also offers a very mixed bag of 'processes' (language transfer, transfer of training), 'strategies' (second language learning, second language communication), and 'overgeneralization', all of which are then also called 'processes'.

Quite apart from the problem of defining and classifying the learner's cognitive strategies, there is a further weakness in most of the attempts made so far to explain errors. This is the tendency to posit one explanation only for a given error. The reason for this tendency may be that many influential studies have been dogmatically oriented, being out to prove or disprove a given hypothesis. Thus Dulay and Burt (1974), for example, do their utmost to argue that certain errors are not interference errors, although they may look like them, but developmental errors (i.e. errors also made in L3-learning), which therefore support their hypothesis that L1-learning and L2-learning. But why do errors have to be one or the
the other type? It is surely more reasonable to assume that an error probably has many sources: a similar structure in the L1 may, for instance, make an overgeneralization of a TL structure more likely.

A third weakness in much EA work is that not enough attention is paid to frequencies. Corder (1975) does mention the importance of this for pedagogical purposes, and it is also stressed by Svartvik (1973); but Ringbom (1976) can still state that the Abo Akademi EA project has a slightly different slant from that of most others, since it is primarily concerned with comparing error frequencies.

In much EA work it seems that one occurrence of a given error is often taken to be as significant as 200 occurrences of another. Types are analyzed, but rarely tokens. Now it is all very well to restrict the analysis to types if the aim is merely to describe the learner's TL system at a given point in the learning process: relative frequencies of occurrence are irrelevant for a description of grammatical competence in the Chomskyan sense. But if the interest is in what native speakers, or learners, do, how they perform, then frequency data cannot be neglected.

Hymes (1971, 1972) made the same point in his paper outlining an approach to the definition and study of communicative competence. We should study not only what is formally possible (i.e. what occurs in the system), but also whether a given form is feasible (in terms of memory and perception limitations, etc.), and appropriate to the situation. We should also be able to state "whether (and to what degree) something is in fact done" (Hymes 1972:281). He adds: "The study of communicative competence cannot be restricted itself to occurrences, but it cannot ignore them." (Hymes 1972:286) In particular, statistical information about what is done is essential if one wants to change what is done.

This view has obvious relevance to the pedagogical, therapy-oriented application of EA.

Selinker's notion of fossilized forms (1972) -- which could be interpreted as relatively permanent errors of high frequency -- captures something of the need to include frequency data.

The importance of statistical information is also stressed by George (1972), who seeks to combine elements from the CA approach and the cognitive approach in his communication-type model. He focusses on the active cognitive processes of selection, organization, and memory storage,
which are inferred to exist in the learner’s ‘black box’. Some causes of error can be traced to faults in the input to the black box, which lead to faulty selection, organization or storage. Input includes not only the textbook, but also the teaching methods, and the whole of the learner’s exposure to the TL. Faulty selection or organization may also be partly due to the natural efficiency-seeking character of the black box, which tends to simplify, to overgeneralize, to reduce redundancy, to secure a maximum return for a given effort (cf. Selinker’s communication strategy). How the learner perceives redundancy in the TL is partly due to how he has been exposed to it, and partly to how he perceives redundancy in his LI. Interference may be from LI, or from already learned parts of the TL, or from other languages known.

George points out that errors are most likely caused by a combination of factors. Interference may combine with redundancy perception and a statistically distorted input, for example, and thus encourage the efficiency-seeking black box to overgeneralize.

I should now like to suggest an extension of this model, which will include insights from both the CA hypothesis and the IL hypothesis, and which will attempt to cope with the various problems discussed above. The model I propose is based on the concept of the linguistic repertoire.

THE STUDY OF THE LEARNER’S LINGUISTIC REPETOIRE

The concept of the linguistic repertoire was defined in sociolinguistic terms by Gumperz (1964) as the totality of linguistic forms regularly employed in the course of socially significant interaction. The same linguistic (or ‘verbal’) repertoire may contain two or more dialects, or even two or more languages. There are social constraints which determine the choice of dialect or language in which a particular utterance is made, but these constraints are seen as operating on a single repertoire.

Within a given speech community the linguistic repertoire may be more or less compartmentalized, less or more fluid. The degree of fluidity or compartmentalization depends partly on the degree of differentiation between the social situations or domains in which each language is used, and partly on the language distance between the languages (or dialects) concerned.

One advantage of the concept of the linguistic repertoire is the
high level of generalization it permits, in that monolingual and multi-
lingual communities may thus both be analysed within the same sociolin-
guistic framework.

Gumperz (1969) extends this concept in his discussion of individual bilingualism. (A given linguistic repertoire may be that of an individual, as well as that of a group.) Still maintaining a sociolinguistic approach, he suggests that in a multilingual community bilinguals shift from language to language in the same way that monolinguals shift from style to style, and that the two languages of bilinguals should be treated as a whole, as one linguistic repertoire.

The fluid-compartmentalized distinction seems related to that between compound and co-ordinate bilingualism (Weinreich 1953). Compound bilingualism, for example, is said to be indicated by a higher degree of L1 interference in L2; this would correspond to a more fluid repertoire.

The acquisition of a first language can be described in these terms as the initial forming of a linguistic repertoire; and I suggest that second language learning may also be described within the same framework. The monolingual learner of an L2 starts with a linguistic repertoire containing only one distinct language (although it may of course, contain more than one style or dialect). The process by which he learns the L2 can be pictured as an enlargement of his repertoire. New items are added to it, from the L2, and by using these he is gradually enabled to widen the scope of his social interaction to include speakers/writers of the L2.

At the beginning of the L2 learning process his repertoire is likely to be fluid, and the 'flow' (or 'interference') will be overwhelmingly towards the L2. As he learns more of the L2, his repertoire continues to enlarge, and it becomes increasingly compartmentalized, as he acquires the ability to use his new L2 in social interaction without recourse to items from his L1.

How can the concept of the linguistic repertoire be applied to EA? It has often been argued (e.g. Corder 1971, Hammarberg 1973) that EA should be concerned not only with the learner's errors but with the whole of his use of the TL. But perhaps we should go further, and study the whole of the learner's linguistic repertoire. Such an extension seems particularly useful for the psycholinguistic application of EA.

If the linguistic repertoire is incorporated into an adaptation of George's model, a picture of the learning of the TL emerges which can also
be helpful in distinguishing types and levels of explanations for errors.

Figure 1 schematizes the growth and use of the linguistic repertoire of a learner (or group of learners) of a given TL. For simplicity, other possible languages in the repertoire are excluded.

The black box now contains both processes and an everchanging product—the linguistic repertoire.

The input may be L1 or TL, and items from this are selected, organized and stored in the expanding repertoire; ideally, they will be stored in the appropriate compartments. Selection and organization processes also occur on the output side, as items from the repertoire are taken and organized into L1 or TL utterances. The learner thus has selection or
choice problems and organization problems (Galanter 1966) at two points.

Output is understood to mean all use of the repertoire, i.e. to include receptive 'use' as well as productive. It is in principle available for direct study, as empirically observable data. The input is observable up to a point, but the processes and the form of the repertoire itself can only be inferred.

If an error is defined as an unwanted form (George 1972) we can apply the term to the output as a whole: L1 utterances, too, may be 'unwanted' in terms e.g. of their social acceptability in a given situation, in terms of their appropriateness. Such a definition also allows us to conceive of errors in realistic, relative terms. In the TL output, for example, a form may be unwanted at one stage of learning, and yet be acceptable at another (e.g. earlier) stage. (Enkvist (1973) has also argued that errors should be defined in relation to specific goals, not absolutely.) We can also include errors of distribution, where the frequencies of items used are 'unwanted' (see e.g. Levenston 1971), and errors concerning the structural and lexical density of the TL output, where the variety of items used is 'unwanted' (see e.g. Linnarud 1976).

Levels and types of explanations for errors can be distinguished by moving step by step up Figure 1. If the TL output contains an unwanted form, the simplest and most superficial explanation is that the learner has selected the wrong item from his repertoire, or has wrongly organized the items selected. Why? Something must be wrong with the repertoire, and the next level of explanation must be found there.

Possible faults in the repertoire are
(a) it is too small - the relevant items or rules are absent;
(b) it contains wrong items or rules - false concepts,
(c) it is statistically distorted;
(d) there is fluidity between the L1 and TL compartments - interference between the two languages. (This assumes that 'fluidity' can apply to individual items or rules, as well as whole languages or dialects.)
(e) there is fluidity within the TL compartment itself - intralingual interference.

Reasons for these faults in the repertoire must be sought first in the processes which go towards making it up: there may have been defects or omissions in the selection or organization of the input, and there
may have been storage (memory, internalization) problems.

Similarly, reasons for these faults go back in turn to the input itself. It may have simply been too small: certain items or rules may not have been present. On the other hand, it may have contained wrong items or rules, it may have been presented or organized in a misleading way, it may have been statistically distorted, etc. Reasons of this kind are normally subsumed under the term 'transfer of training'.

The relevance of Contrastive Analysis is also evident here: the way in which the L1 input has been processed will obviously affect the selection, organization and storage of the TL input.

What have been termed 'strategies' are now seen to operate at different stages, directed at different objects: there are the strategies involved in selecting and organizing the input, those involved in selecting and organizing items from the repertoire for the output, and those motivated by the act of communication as a whole, by the effort to transmit meaning.

Underlying all these explanations, at all stages, is the general tendency of the black box to be efficiency-seeking, to reduce redundancy, to simplify, etc. Thus fluidity will be more likely where it produces greater simplicity and efficiency in the repertoire; items will tend not to be stored if they are seen to be unduly complicated and to offer inadequate compensation in terms of increased output potential; input will be organized as simply as possible; and so on.

It therefore becomes clear that an error cannot be given one explanation only, but must be traced back through the various stages in the operations of the black box. And there are of course many other factors which may be relevant to the explanation of a given error, and which are not shown in this model: factors such as the age and sex of the learner, the time spent learning, the character of his exposure to the TL, the learner's attitude to the TL, his motivation, etc. In theory one would have to analyse the whole of the learner's previous experience of life.

At a practical level, though, it is often possible to identify a major cause for a given error. Further testing, careful elicitation and follow-up, such as the lateralization techniques discussed by Kellerman (1974), can sometimes pinpoint specific defects in the black box. If a learner cannot correct an error when asked to do so, for example, we
may assume that the relevant rule or item is not present in the repertoire, at least has not been sufficiently internalized. Perhaps the most useful practical guide is the learner's own perception of the error (see Kellerman 1974, Tran-Thi-Chau 1975), rather than what the analyst thinks the primary explanation is. Further experimental work is certainly needed, e.g. on how variations in the presentation of the input affect the output.

From the pedagogical point of view the best explanation is obviously the one which, when made use of in syllabus planning or remedial teaching, most reduces the frequency of the error. Here again, experimentation is necessary.

For pedagogical purposes we also need to evaluate errors in terms of their degree of seriousness or degree of deviance. This is a complex problem, and I do not intend to discuss it here (see e.g. Nickel 1973, Johansson 1973 and 1975, Chesterman 1977). It has been realised that linguistic criteria of one kind or another often conflict with communication criteria; and other factors, such as the stage of learning reached, complicate the issue still further. In the present stage of the art the safest solution would seem to be that any scale of deviance used should be as narrow as possible; the more degrees of seriousness that are recognized, the greater the danger of a loss of validity in the grading scale. Much work remains to be done here.

To conclude: despite the problems involved in trying to explain (and evaluate) errors, EA does have a claim to be psycholinguistically and pedagogically useful. I believe this claim will be all the more substantial if the discipline can include the kind of sociolinguistic dimension suggested here.
BIBLIOGRAPHY


Corder, S.P. 1971. Idiosyncratic dialects and error analysis, IRAL 9, 147-60.


