

From:  
 Ellis, R. (1994): *The Study of Second Language Acquisition*.  
 Oxford: OUP.

An error can be defined as a deviation from the norms of the target language. This definition raises a number of questions, however. First, there is the question regarding which variety of the target language should serve as the norm. The general practice, especially where classroom learners are concerned, is to select the standard written dialect as the norm. This, of course, is fundamentally wrong if the goal is to describe learners' oral production. Nor is it always possible to adopt the standard spoken variety as the norm. Some learners are exposed to varieties of the language which differ from the standard dialect. For example, in comparison with the norms of British or American standard written English the utterance

\*She coped up with her problem very well.

is erroneous, but in comparison with norms of educated Zambian English such an utterance can be considered correct.<sup>5</sup>

A second question concerns the distinction between *errors* and *mistakes* (Corder 1967). An error (in this technical sense) takes place when the deviation arises as a result of lack of knowledge. It represents a lack of competence (see Chapter 1, page 12). A mistake occurs when learners fail to perform their competence. That is, it is the result of processing problems that prevent learners from accessing their knowledge of a target language rule and cause them to fall back on some alternative, non-standard rule that they find easier to access. Mistakes, then, are performance phenomena and are, of course, regular features of native-speaker speech, reflecting processing failures that arise as a result of competing plans, memory limitations, and lack of automaticity. Corder argues that the EA should be restricted to the study of errors (i.e. mistakes should be eliminated from the analysis). However, apart from the problems of identification that this distinction raises, it also assumes that competence is homogeneous rather than variable. Thus, if learners sometimes use a correct target form and sometimes an incorrect, non-target form, it cannot necessarily be concluded that the learner 'knows' the target form and that the use of the non-target form represents a mistake. It is possible that the learner's knowledge of the target form is only partial; the learner may not have learnt all the contexts in which the form in question can be used. For example, a learner may have no difficulty in using the target language form in some linguistic contexts:

My sisters are older than me.

but produce an error in others:

\*My three sister are older than me.

In this early period, the study of learner errors largely ignored the problem of variability in learner language (see Chapter 4).

### Identification of errors

Once a corpus of learner language has been collected, the errors in the corpus have to be identified. It is necessary to decide, therefore, what constitutes an 'error' and to establish a procedure for recognizing one.

A third question concerns whether the error is overt or covert (Corder 1971a). An overt error is easy to identify because there is a clear deviation in form, as when a learner says:

\*I runned all the way.

A covert error occurs in utterances that are superficially well-formed but which do not mean what the learner intended them to mean. For example, the utterance (from Corder 1971a):

\*It was stopped.

is apparently grammatical until it becomes clear that 'it' refers to 'the wind'. Furthermore, a superficially correct utterance may only be correct by chance. For example, the learner may manifest target-like control of negative constructions in ready-made chunks such as 'I don't know' but fail to do so in 'created' utterances (i.e. utterances that are constructed on the basis of rules the learner has internalized). The existence of covert errors led Corder to argue that 'every sentence is to be regarded as idiosyncratic until shown to be otherwise' (page 21).

A fourth question concerns whether the analysis should examine only deviations in correctness or also deviations in appropriateness. The former involves rules of usage and is illustrated in the two examples above. The latter involves rules of language use. For example, a learner who invites a relative stranger by saying 'I want you to come to the cinema with me' has succeeded in using the code correctly but has failed to use it appropriately. In general, EA has attended to 'breaches of the code' and ignored 'misuse of the code' (Corder 1974: 124), but more recently attention has been paid to the latter (Thomas 1983). Errors of this pragmatic kind will be considered in Chapter 5.

These various distinctions are indicative of the kinds of problems which analysts face in recognizing errors. To overcome them, Corder (1971a; 1974) proposes an elaborate procedure for identifying errors, which is shown in Figure 2.1. This procedure acknowledges the importance of 'interpretation' and distinguishes three types: normal, authoritative, and plausible. A normal interpretation occurs when the analyst is able to assign a meaning to an utterance on the basis of the rules of the target language. In such cases, the utterance is 'not apparently erroneous', although it may still only be right 'by chance'. An authoritative interpretation involves asking the learner (if available) to say what the utterance means and, by so doing, to make an 'authoritative reconstruction'. A plausible interpretation can be obtained by referring to the context in which the utterance was produced or by translating the sentence literally into the learner's L1.

There are a number of major methodological problems with the procedures used in error identification, some of which we have already noted. Corder's (1967) distinction between errors and mistakes is not easy to put

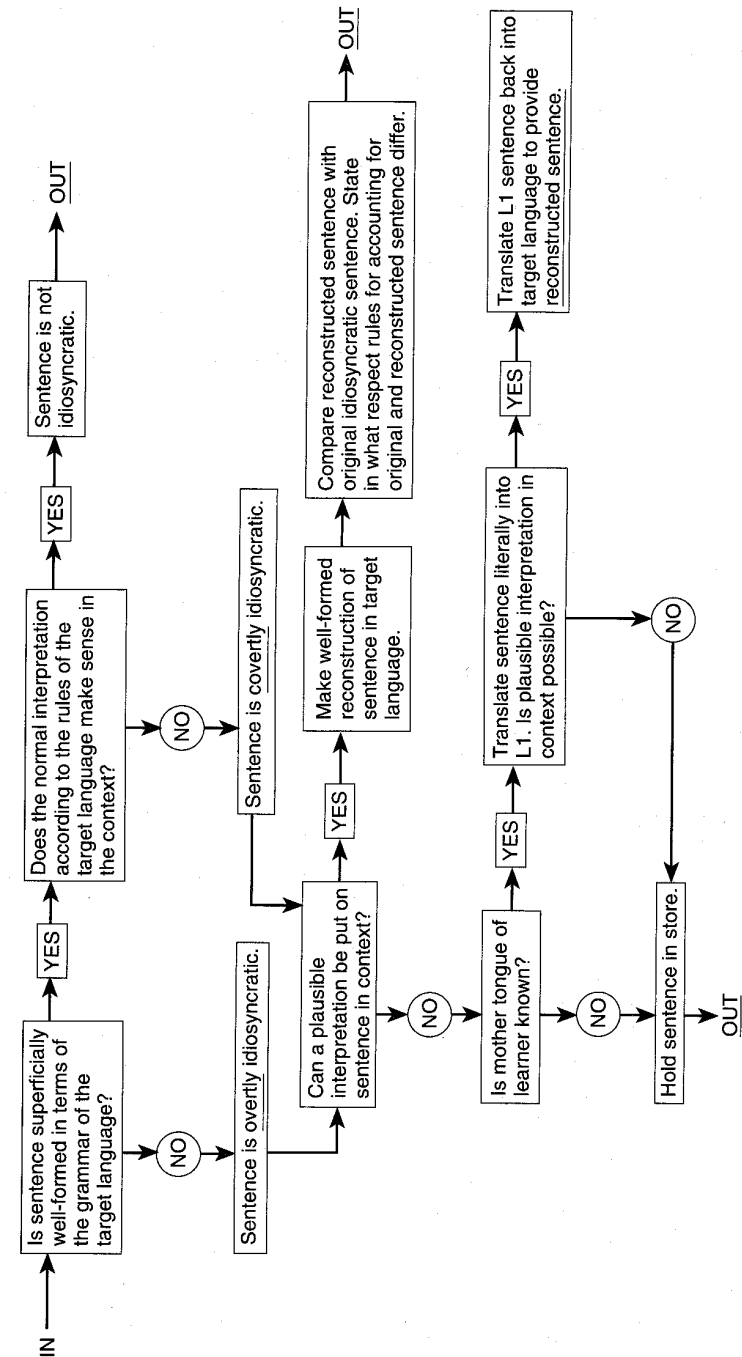


Figure 2.1: Algorithm for providing for description of idiosyncratic dialects (from Corder 1981a)

into practice, even if the learner is available to provide an 'authoritative interpretation'. In addition, the distinction does not take account of the possibility that learners' knowledge is variable. Also, it is not at all clear whether Corder's suggestions for identifying covert errors will work. Reliance on the learner as an informant has been criticized on the grounds that retrospective accounts of intended meaning are often not reliable (Van Els et al. 1984). Also, such a procedure assumes that learners possess the necessary *metalingual knowledge* to talk about their own performance—an assumption that may not be justified in the case of children and some adult learners. Not all researchers have found problems with identifying errors, however. Duskova (1969), in a study considered in the next section, discusses it at some length and concludes that 'the number of cases in which it was hard to decide whether an error had been made . . . did not exceed 4 per cent of all the errors examined'. Unfortunately, Duskova does not provide inter-rater reliability measures for the errors identified in her sample—a failing common to most of the early studies.<sup>6</sup>

### Description of errors

The description of learner errors involves a comparison of the learner's idiosyncratic utterances with a reconstruction of those utterances in the target language. It requires, therefore, attention to the surface properties of the learners' utterances (i.e. it does not attempt, at this stage, to identify the sources of the errors). Some researchers have felt the need to maintain a clear distinction between the description and explanation of errors. Dulay, Burt, and Krashen (1982), for example, argue the need for descriptive taxonomies of errors that focus only on observable, surface features of errors, as a basis for subsequent explanation.

Perhaps the simplest type of descriptive taxonomy is one based on linguistic categories. This type is closely associated with a traditional EA undertaken for pedagogic purposes, as the linguistic categories can be chosen to correspond closely to those found in structural syllabuses and language text books. An example can be found in Burt and Kiparsky's *The Gooficon: A Repair Manual for English* (1972). This identifies a number of general linguistic categories (for example, the skeleton of English clauses, the auxiliary system, passive sentences, temporal conjunctions, and sentential complements). Each general category is then broken down into further levels of subcategories. For example, the auxiliary system is subdivided into 'do', 'have and be', modals, and mismatching auxiliaries in tag questions, while errors in the use of 'do' are classified according to whether they involve over-use in questions and negatives, underuse in questions or overuse in affirmative sentences. Politzer and Ramirez (1973) begin with even more general categories: morphology, syntax, and vocabulary. Such taxonomies allow for both a detailed description of specific errors and also for a quantification of a corpus of errors.

The 1960s saw a number of studies which provided descriptions of the different kinds of linguistic errors produced by learners. Richards (1971b), in a paper designed to challenge the widely held belief that learner errors were the result of L1 interference, provided a taxonomy of different categories of linguistic error based on a number of previous studies. He examined errors made by learners from different language backgrounds (Japanese, Chinese, Burmese, French, Czech, Polish, Tagalog, Maori, Maltese, and the major Indian and West African languages) and illustrated the different kinds of errors relating to the production and distribution of verb groups, prepositions, articles, and the use of questions. However, he made no attempt to quantify the errors. Nor do we know to what extent his linguistic categories accounted for all the errors he examined. Duskova (1969)—one of the studies Richards drew on—is better in this respect. She identified a total of 1007 errors in the written work of 50 Czech learners of English, who were postgraduate students studying science. She found 756 'recurrent systemic errors' and 251 'nonce errors' (i.e. errors that occurred once only). Errors in articles were most common (260), followed by errors in lexis (233) and morphology (180). In comparison, there were only 54 errors in syntax and 31 in word order. Duskova noted, however, that the frequency of the errors did not necessarily reflect the level of difficulty the learners experienced with different linguistic features, as some features (such as articles) were attempted more often than others (for example, adverbs). Duskova also noted that although she had few difficulties in assigning errors to general linguistic categories such as 'word order', it often proved very difficult to classify them accurately into subcategories.

These studies were cross-sectional in design. Of greater interest for SLA research are longitudinal studies of learners' errors as these can show in what areas of language errors persist over time. Chamot's (1978; 1979) study of the acquisition of English by a bilingual French/Spanish boy is interesting in this respect. She found that the main linguistic problem areas were omission of constituents, verb forms, sentence formation, articles, and prepositions. In some of these (for example, omission of constituents) the number of errors reduced sharply over a 44-month period, while in others (for example, question formation) little improvement was evident. In all the areas, however, there was considerable fluctuation in error frequency throughout the period. Chamot's study suggests that it may be difficult to provide a satisfactory description of learners' L2 development by quantifying the types of errors they make.

An alternative to a linguistic classification of errors is to use a surface strategy taxonomy. This 'highlights the ways surface structures are altered' (Dulay, Burt, and Krashen 1982: 150) by means of such operations as omissions, additions and regularizations. Table 2.2 provides a part of the total taxonomy together with examples of the categories. Dulay, Burt, and Krashen claim that such an approach is promising because it provides an

indication of the cognitive processes that underlie the learner's reconstruction of the L2. This seems a doubtful claim, however, as it presupposes that learners operate on the surface structures of the target language rather than create their own, unique structures. If a surface strategy taxonomy does not represent mental processes, it is not clear what value it has. This may account for why there have been few attempts to describe learner errors using such a taxonomy.

Category	Description	Example
Omissions	The absence of an item that must appear in a well-formed utterance.	She sleeping.
Additions	The presence of an item that must not appear in well-formed utterances.	We didn't went there.
Misinformations	The use of the wrong form of the morpheme or structure.	The dog ated the chicken.
Misorderings	The incorrect placement of a morpheme or group of morphemes in an utterance.	What daddy is doing?

Table 2.2: *A surface strategy taxonomy of errors (categories and examples taken from Dulay, Burt, and Krashen 1982)*

Although linguistic and surface strategy taxonomies of errors may have a pedagogic application (for example, by demonstrating which errors are the most frequent and, therefore, most in need of attention), in general they shed little light on how learners learn an L2. Corder's (1974) framework for describing errors is more promising in this respect. He distinguishes three types of error according to their systematicity:

- 1 Presystematic errors occur when the learner is unaware of the existence of a particular rule in the target language. These are random.
- 2 Systematic errors occur when the learner has discovered a rule but it is the wrong one.
- 3 Postsystematic errors occur when the learner knows the correct target language rule but uses it inconsistently (i.e. makes a mistake).

In order to identify these different kinds of errors, however, it is necessary to interview the learner. Thus, type (1) occurs when the learner cannot give any account of why a particular form is chosen, type (2) occurs when the learner is unable to correct the errors but can explain the mistaken rule used, and type (3) occurs when the learner can explain the target-language rule that is normally used. Such a taxonomy, therefore, requires that the researcher has access to the learners and that the learners are capable of providing explanations for their L2 behaviour. For this reason it may prove difficult to operate.

The description of errors, like their identification, is problematic. Even if the error itself can be easily identified, it is often problematic to determine what the error consists of. If a learner produces the following sentence:

\*My name Alberto.

there is no difficulty in reconstructing the target-language version:

My name is Alberto.

and so establishing that copula 'is' has been omitted. But in many cases—even with sentences that are overtly idiosyncratic—the reconstruction of the target language version—and, therefore, its description—is problematic. For example, if a learner produces the following sentence:

\*I am worried in my mind.

it is not clear what constitutes the best reconstruction. One possibility is 'I am feeling worried.' Another is 'I have a problem on my mind.'

Even if the learner is available for consultation, it may not be possible to choose between these two reconstructions. But the description of the error will obviously vary according to which reconstruction is finally chosen. The reconstruction of covertly idiosyncratic sentences will prove even more difficult.

Another problem concerns the failure to quantify the different types of errors that have been identified and described. Many of the EA studies that have been conducted have been very informal—perhaps as a result of the kinds of problems discussed above. In some studies error frequencies are not given at all (for example, Jain 1974; Richards 1971b), while in others only absolute frequencies are given (for example, Duskova 1969). But as Schachter and Celce-Murcia (1977) point out, to say anything worthwhile about error frequency we need to know the number of times it would be possible for learners to have committed different errors. In other words, relative rather than absolute frequencies are needed.

### Explanation of errors

Assuming that it is possible to identify and describe errors, the next step is to try to explain them. Explanation is concerned with establishing the source of the error, i.e. accounting for why it was made. This stage is the most important for SLA research as it involves an attempt to establish the processes responsible for L2 acquisition.

As Taylor (1986) points out, the error source may be psycholinguistic, sociolinguistic, epistemic, or may reside in the discourse structure. Psycholinguistic sources concern the nature of the L2 knowledge system and the difficulties learners have in using it in production. Sociolinguistic sources involve such matters as the learners' ability to adjust their language in

accordance with the social context. Epistemic sources concern the learners' lack of world knowledge, while discourse sources involve problems in the organization of information into a coherent 'text'. In general, however, SLA research has attended only to the first of these. As Abbott puts it: 'The aim of any EA is to provide a psychological explanation' (1980: 124). Figure 2.2 plots the different psycholinguistic sources to be discussed.

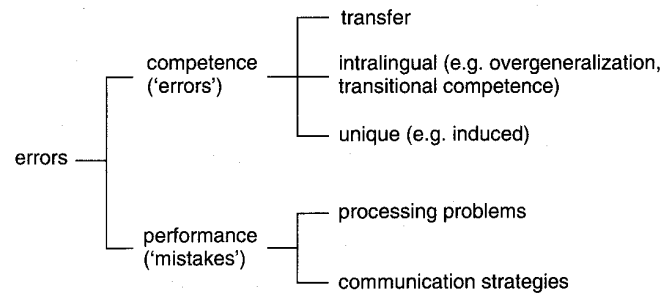


Figure 2.2: *Psycholinguistic sources of errors*

The distinction between 'errors' and 'mistakes', which has already been discussed with regard to the identification stage of EA, is also relevant in explaining deviations—demonstrating the interdependence of these two steps in EA. Any deviation from target-language norms may reflect either a problem in performance or in competence.<sup>7</sup> It is helpful to recognize two different kinds of performance mistake: those that result from processing problems of various kinds, and those that result from such strategies as circumlocution and paraphrase, which a learner uses to overcome lack of knowledge. The latter are known as communication strategies and will be discussed in some detail in Chapter 9. As we have already seen, it is competence errors that have been considered central to the study of L2 acquisition.

A number of different sources or causes of competence errors have been identified. Richards (1971b) distinguishes three:

- 1 Interference errors occur as a result of 'the use of elements from one language while speaking another.' An example might be when a German learner of L2 English says \*'I go not' because the equivalent sentence in German is 'Ich gehe nicht'.
- 2 Intralingual errors 'reflect the general characteristics of rule learning such as faulty generalization, incomplete application of rules and failure to learn conditions under which rules apply'.
- 3 Developmental errors occur when the learner attempts to build up hypotheses about the target language on the basis of limited experience.

However, Schachter and Celce-Murcia (1977) find the distinction between intralingual and developmental errors 'curious', and most researchers have

operated with a general distinction between transfer errors (Richards' category (1)) and intralingual errors (an amalgam of Richards' (2) and (3)).

Transfer errors can be further subdivided. Lott (1983), for instance, distinguishes three categories:

- 1 'Overextension of analogy' occurs when the learner misuses an item because it shares features with an item in the L1 (for example, Italian learners use 'process' to mean 'trial' because Italian 'processo' has this meaning).
- 2 'Transfer of structure' arises when the learner utilizes some L1 feature (phonological, lexical, grammatical, or pragmatic) rather than that of the target language. This is what is generally understood as 'transfer'.
- 3 'Interlingual/intralingual errors' arise when a particular distinction does not exist in the L1 (for example, the use of 'make' instead of 'do' by Italian learners because the 'make/do' distinction is non-existent in Italian).

It is this third category that has caused so many of the problems in determining whether an error is transfer or intralingual.

Intralingual errors are also often further subdivided. Thus, Richards (1971b) distinguishes the following:

- 1 Overgeneralization errors arise when the learner creates a deviant structure on the basis of other structures in the target language. It generally involves the creation of one deviant structure in place of two target language structures (for example, \*'He can sings' where English allows 'He can sing' and 'He sings').
- 2 Ignorance of rule restrictions involves the application of rules to contexts where they do not apply. An example is \*'He made me to rest' through extension of the pattern found with the majority of verbs that take infinitival complements (for example, 'He asked/wanted/invited me to go').
- 3 Incomplete application of rules involves a failure to fully develop a structure. Thus learners of L2 English have been observed to use declarative word order in questions (for example, \*'You like to sing?') in place of interrogative word order (for example, 'Do you like to sing?'). This type of intralingual error corresponds to what is often referred to as an error of *transitional competence* (Richards 1971a).
- 4 False concepts hypothesized (i.e. the learner fails to comprehend fully) arise when the learner does not fully comprehend a distinction in the target language—for example, the use of 'was' as a marker of past tense in \*'One day it was happened'.

It is not easy to distinguish transfer and intralingual errors, and even more difficult to identify the different types of intralingual errors that Richards describes. In an attempt to deal with the problem of identifying sources, Dulay

and Burt (1974b) classified the errors they collected into three broad categories:

- 1 Developmental (i.e. those errors that are similar to L1 acquisition)
- 2 Interference (i.e. those errors that reflect the structure of the L1)
- 3 Unique (i.e. those errors that are neither developmental nor interference).

Dulay and Burt's research has often been criticized on the grounds that reliable classification of errors in terms of these categories is still not possible. However, it might be argued that by using L1 acquisition errors as a baseline they at least provide an operational procedure for establishing which errors are intralingual.

It is customary to distinguish another general source of errors. *Induced errors* occur when learners are led to make errors by the nature of the instruction they have received. Stenson (1974) provides a number of examples of such errors in the classroom speech of Tunisian learners of English. Faulty explanation of grammatical points can give rise to errors (for example, the use of 'any' to mean 'none' when the students were told that 'any' has a negative meaning). Drills performed without consideration for meaning can also result in error. Svartvik (1973b) suggests that overdrilling may be one of the reasons why Swedish learners of L2 English overuse infinitival complements (for example, \*'He proposed her to stay'). Stenson argues that such errors are not systematic and therefore do not reflect competence. However, there are probably cases when learners do internalize faulty rules derived from instruction and in such cases the resulting errors will reflect their competence. Instruction may constitute one source of what Dulay and Burt call 'unique errors'.

The bulk of the empirical work in SLA has focused on determining what proportion of the total errors in a corpus are transfer as opposed to intralingual. This is what motivated such studies as Richards (1971b) and Dulay and Burt (1974b). The issue that these researchers tackled was the competing claims of a behaviourist, habit-formation account of L2 acquisition and a mentalist, creative-construction account. According to behaviourist accounts, errors were viewed as the result of the negative transfer of L1 habits. According to mentalist accounts, errors were predicted to be similar to those found in L1 acquisition because learners actively construct the grammar of an L2 as they progress (i.e. they are intralingual). This issue dominated early work in SLA. It should be noted, however, that subsequently researchers have come to recognize that the correlation between behaviourism and transfer errors on the one hand and mentalism and intralingual errors on the other is simplistic and misleading. Transfer is now treated as a mental process in its own right (see Chapter 8).

A good example of the kind of 'proportion study' that investigated this behaviourist/mentalistic question is Dulay and Burt (1974b). Five hundred and thirteen unambiguous errors produced by Spanish children acquiring L2

English were extracted from a corpus of speech collected by means of the Bilingual Syntax Measure (see page 50 in this chapter). These errors occurred in six syntactic structures which differed in English and Spanish. The errors were classified as 'developmental', 'interference', or 'unique', and detailed results for each structure were provided. In each case the developmental errors far outweighed the interference errors. For example, for the structure NP + V + Pronoun (for example, 'The dog ate it'), which is realized as NP + Pronoun + V in Spanish (for example, 'El perro se lo comio'), there were 93 developmental errors and no interference errors. Overall less than 5 per cent of the total errors were attributed to interference. This led Dulay and Burt to propose that L1 and L2 acquisition were very similar (the L2=L1 Hypothesis). It should be noted, however, that other studies (for example, George 1972 and Flick 1980) found a much higher proportion of transfer errors (see Chapter 8).

One reason for the discrepancy in research findings has been the problem of error classification. As Flick notes:

The assignment of a particular error to such categories as 'transfer', 'overgeneralization' or 'ambiguous' has been largely an arbitrary matter, subject to the individual biases and point of view of the researcher (Flick 1979: 60).

To solve this problem, Flick proposed the use of factor analysis. This is a statistical procedure which uses the patterns of correlation that exist among scores on different variables to identify underlying factors. Flick reported on the preliminary use of this procedure to identify clusters of errors in the speech elicited from 20 adult Spanish learners of L2 English by means of an oral translation task. Five factors emerged from the analysis:

- 1 transfer accounted for 34 per cent of the variance in the scores assigned to each learner
- 2 performance (i.e. constraints imposed by processing load) for 23 per cent
- 3 simplifications of function words, such as errors of omission, for 17 per cent
- 4 overgeneralization for 16 per cent
- 5 pronominal reference (for example, wrong choice of pronoun) for 11 per cent.

Flick's study represents an interesting attempt to increase the rigour of error explanation, but it is open to a number of objections. For example, the five categories identified constitute a very mixed bag: while some are explanatory in nature, others seem more descriptive.

It is difficult to synthesize the results of attempts to explain errors in learner language, but the following appear to be some of the main findings:

- 1 A large number—and in some cases perhaps most—of the errors that learners produce are intralingual in origin rather than transfer.

However, the precise proportion of the kinds of error varies considerably from study to study.

- 2 According to Taylor (1975), learners at an elementary level produced more transfer errors than learners at an intermediate or advanced level. Conversely, he found that learners at an intermediate or advanced level produced more intralingual errors (for example, overgeneralization) than learners at an elementary level. However, as we will see in Chapter 8, other researchers (for example, Kellerman 1983) have challenged the view that transfer is more prevalent in beginners.
- 3 The proportion of transfer and intralingual errors varies in accordance with the task used to elicit samples of learner language. Thus, translation tasks tend to result in more transfer errors than tasks that call for free composition (Lococo 1976).
- 4 Transfer errors are more common in the phonological and lexical levels of language than in the grammatical level. Also some areas of grammar acquisition are more likely to be influenced by the learners' L1 than others. Grauberg (1971) found that interference accounted for 25 per cent of the lexical errors produced by adult German learners of L2 English, 10 per cent of their syntactic errors, and none of their morphological errors.
- 5 Transfer errors are more common in adult learners than in child learners. For example, White (1977) found that 21 per cent of the errors made by adult Spanish learners of English were transfer. White used the same instrument to collect data as in the Dulay and Burt (1974b) study of Spanish children referred to above, so this study is directly comparable.
- 6 Errors can have more than one source. For example, the 'no' + verb error (as in \*'No look my card') is universal, suggesting an intralingual explanation, but Spanish learners of L2 English have been noted to make this error more frequently and for a longer period of time, suggesting that the L1 pattern for negatives (*no* + verb) is also having an influence.

Again, it is important to recognize that the concept of 'transfer', upon which many of the early EA studies were based, was simplistic. Transfer is, in fact, a very complex notion which is best understood in terms of cognitive rather than behaviourist models of learning.

It should be clear from this account of the explanation stage of EA that problems abound. In particular, the concepts of transfer and intralingual error were often not operationalized with sufficient rigour. Where one researcher identified the source of an error as transfer, another researcher identified the source of the same error as intralingual. For example, Duskova (1969) interpreted article deletion in Czech learners of English as interference, while Dulay and Burt (1974b) interpreted the same error in Spanish children learning English as intralingual. Schachter and Celce-Murcia (1977) argue that a large number of learners' errors are ambiguous with regard to source and that 'one must be extremely cautious when claiming to have

identified the cause of any given error type'. Such caution has not always been exercised, however. If clear explanatory statements about errors are often not possible, the value of EA as a tool for investigating L2 acquisition is thrown into question.

### Evaluating errors

Whereas all the preceding stages of EA have involved an examination of errors from the point of view of the learner who makes them, *error evaluation* involves a consideration of the effect that errors have on the person(s) addressed. This effect can be gauged either in terms of the addressee's comprehension of the learner's meaning or in terms of the addressee's affective response to the errors. Error evaluation studies proliferated in the late 1970s and in the 1980s, motivated quite explicitly by a desire to improve language pedagogy.<sup>8</sup> The studies surveyed in Table 2.3 constitute only a part of the total. Ludwig (1982) provides a survey of twelve early studies.

The design of error evaluation studies involves decisions on who the addressees (i.e. the judges) will be, what errors they will be asked to judge, and how they will be asked to judge them. The judges can vary according to whether they are native speakers (NS) or non-native speakers (NNS), and also according to whether they are 'expert' (i.e. language teachers) or 'non-expert'. The errors they have been asked to judge cover semantic or lexical aspects of English, different grammatical features, and spelling. The instruments used to elicit judgements vary in a number of ways. In most cases they consist of decontextualized lists of sentences containing either one or several errors. These sentences are usually taken from actual samples of learner language (mainly written compositions) but they are sometimes contrived. In some studies the sentences are contextualized. The errors can be presented orally but are usually presented in writing. The judges may be asked to evaluate the 'comprehensibility' of the sentences containing the errors, the 'seriousness' or the 'naturalness' of errors, or the degree of 'irritation' they arouse. Sometimes they may be asked to correct the errors and to give reasons for why they judged some errors as especially problematic. In some studies the judges' comprehension of the erroneous sentences is also tested.

Error evaluation studies have addressed three main research questions: (1) Are some errors judged to be more problematic than others? (2) Are there differences in the evaluations made by NS and NNS? and (3) What criteria do judges use in evaluating learners' errors? We will briefly consider the main findings on each of these issues.

NS judges tend to judge lexical errors as more serious than grammatical errors (for example, Burt 1975; Tomiyana 1980; Khalil 1985). They also tend to judge global grammatical errors as more likely to interfere with comprehension than local errors, although as Santos (1987) points out, there have been conflicting results on this point. Burt defines *global errors* as errors that

Study	Subjects	Measures	Procedure	Main Results
Burt 1975	Non-expert NS.	Partially corrected versions of 300 sentences containing multiple errors.	Subjects were asked to judge comprehensibility of different corrected versions.	Subjects found versions in which 'global' errors had been corrected more comprehensible than versions in which 'local' errors had been corrected.
Albrechtsen, Henriksen, and Faerch 1980	120 non-expert adult NNS (e.g. hotel workers in UK; 180 British sixth-formers).	Samples of oral language taken from Grade 10 Danish learners of L2 English; the samples varied with regard to error density.	Subjects listened to tapes and rated each sample using bipolar adjective scales (e.g. easy to understand-difficult to understand).	The oral texts containing few errors (syntactic as well as lexical) and few communication strategies (CSs) received positive evaluations. Frequent use of CSs had greater negative effect than number of errors.
Tomiyana 1980	NS graduates in Education and Language departments	2 constructed passages (200 words in length), designed to include 7 instances of six kinds of errors involving articles and sentence connectors.	Subjects were asked to read the passages and correct the errors and to rate likely academic achievement of the writer of the passage.	Subjects corrected insertion errors more accurately than omission or wrong choice errors and article errors more accurately than connector errors. Errors in articles not perceived as so damaging to academic success as errors in connectors.
Chastain 1981	27 native Spanish speakers; undergraduates at university in Spain.	10 paragraphs written by American university students of Spanish.	Subjects were asked to read the passages, underline each error, and then evaluate it as comprehensible and acceptable, comprehensible but not acceptable, or not comprehensible.	The overall seriousness of the different errors was (1) word errors in noun phrases (most serious); (2) form errors in verb phrases; (3) word errors in verb phrases; and (4) form errors in noun phrases (least serious).
Hughes and Lascaratou 1982	10 Greek-speaking teachers; 10 NS teachers; 10 non-expert NS.	32 sentences containing errors in 8 categories (vocabulary, grammar, and spelling) taken from learners' compositions; 10 error-free sentences.	Subjects were asked to underline errors, write correct versions, and judge their seriousness. Also asked to give reason if error judged 'very serious'.	NNS judged errors overall as more serious than NS. NNS more lenient on spelling errors. NNS judged according to whether error constituted infringement of 'basic' rule; NS judged according to intelligibility. Evaluation of <small>seriousness</small> <small>more lenient</small>
Study	Subjects	Measures	Procedure	Main Results
Davies 1983	43 Moroccan teachers of English; 43 non-expert NS.	82 contrived sentences (some correct) containing typical errors of Moroccan secondary school students.	Subjects were asked to rate seriousness of the errors and to add comments.	NS more lenient than NNS. NNS very hard on morphological and tense choice errors, but NNS less hard on obvious transfer errors and 'global' errors.
Vann et al. 1984	319 faculty members of US university.	24 sentences containing multiple errors and 12 containing single errors; errors in 12 error categories.	Subjects were asked to judge acceptability of sentences.	Subjects more prepared to accept errors of kind made by NS students (e.g. in spelling), but less likely to accept 'global' errors (e.g. word order).
Khalil 1985	240 American undergraduate NS.	20 grammatically deviant and 12 semantically deviant sentences taken from compositions by Arab 1st year students, some contextualized and some not.	Subjects were asked to judge intelligibility and naturalness of sentences. They were also tested on the ability to understand the sentences' meaning.	Semantic errors were judged less intelligible and found less comprehensible than grammatical errors. Contextualization of sentences did not improve intelligibility.
Sheorey 1986	64 NS teachers and 34 NNS (Indian) teachers.	20 sentences reflecting 8 categories of error taken from compositions written by foreign students at university in USA.	Subjects were asked to judge seriousness of errors.	NNS judged errors overall more serious than NS. NS judged lexical errors more serious than NNS. NNS judged errors in tense, agreement, prepositions, question formation, and spelling more serious than NS.
Santos 1987	40 university professors in physical sciences.	4 written compositions; in 2 of them 5 errors of the marked-to-unmarked kind (e.g. 'a' instead of 'an') were inserted and in the other 2, 5 errors of the unmarked-to-marked kind (e.g. 'an' instead of 'a') were inserted.	Subjects were asked to rank each composition and to underline each error and to assess the degree of irritation it aroused.	No significant difference in rankings of the compositions. Overall, unmarked-to-marked errors were found to be more irritating than marked-to-unmarked errors. Syntactic errors were also found more irritating than morphological errors.

Table 2.3: *A summary of selected error evaluation studies*



affect overall sentence organization. Examples are wrong word order, missing or wrongly placed sentence connectors, and syntactic overgeneralizations. *Local errors* are errors that affect single elements in a sentence (for example, errors in morphology or grammatical functors). NS judges may also be influenced by markedness factors. Santos' study, for example, lends some credence to the idea that errors involving the substitution of marked for unmarked forms (for example, 'an book' for 'a book') are judged more severely than errors in which unmarked forms replace marked forms (for example, 'a apple' for 'an apple'). NS judges also find it easier to deal with insertion than with omission or wrong choice errors (Tomiya 1980). It should be noted, though, that there can be considerable variation in the judgements of native speakers. Thus Vann, Meyer, and Lorenz (1984) found that some academic faculty members were inclined to view all errors as equally serious—'an error is an error'. Also, attempts to identify a hierarchy of errors according to their effect on intelligibility have not proved successful.

There are clear differences in the judgements made by NS and NNS. Overall, NNS are much more severe (James 1977; Hughes and Lascaratou 1982; Davies 1983; Sheorey 1986). NNS judges seem to be especially hard on morphological and functor errors in comparison to NS judges. However, they tend to evaluate lexical and global errors less severely than NS judges.

Judges appear to use different criteria in assessing *error gravity*. Khalil (1985) identifies three general criteria: intelligibility, acceptability, and irritation. Intelligibility concerns the extent to which sentences containing different kinds of error can be comprehended. Acceptability is a rather vague criterion, involving judgements of the seriousness of an error. Irritation concerns the emotional response of an addressee but is also related to the frequency of errors. Albrechtsen, Henriksen, and Færch (1980), who had NS judges rate the errors made by Danish learners of English in oral interviews, found that 'all errors are equally irritating . . . irritation is directly predictable from the number of errors regardless of the error type or other linguistic aspects' (1980: 394).

Obviously NS and NNS judges vary in the criteria they use. NS judges appear to be more concerned with the effect that an error has on their comprehension, whereas NNS judges are more influenced by their ideas of what constitute the 'basic' rules of the target language (Hughes and Lascaratou 1982). However, Davies (1983) points out that NNS judgements will be influenced by a number of factors relating to the particular context in which they operate. Thus NNS teachers will be influenced by their background knowledge of the syllabus and text book the learners are following and by explicit knowledge of their L1. Transfer errors are viewed leniently, but errors in grammatical structures that have already been taught will be seen as more serious. Davies makes the general point that 'any evaluation will be coloured

by the particular viewpoint from which it was carried out, and this may not be consistent with evaluations made from other viewpoints' (1983: 310).

As we noted in the introduction to this chapter, EA studies have often been pedagogically motivated. They have sought to identify criteria for establishing error gravity so that teachers can be guided in what errors to pay more attention to. The general conclusion is that teachers should attend most carefully to errors that interfere with communication (i.e. semantic and global grammatical errors).<sup>9</sup> Johansson (1973) suggests that errors should be evaluated by first asking whether they are comprehensible, and second whether they cause irritation. Other, secondary factors—the frequency and generality of the feature involved—also need to be considered. In this way, Johansson constructs a hierarchy of errors. However, he acknowledges that 'it is not possible to illustrate the scale of errors at the present time since there is no available information concerning the degree of comprehensibility/irritation caused by different errors' (1973: 109). Although there has been considerable research since, there is still insufficient evidence to support a definite scale for evaluating errors.

Like other aspects of EA, the evaluation of learner error poses a number of problems. It is not at all clear what criteria judges use when asked to assess the 'seriousness', 'intelligibility', or 'acceptability' of an error. As we have seen, error evaluation is influenced by the context in which the errors occurred. Thus, the same error may be evaluated very differently depending on who made it and where, when and how it was made. The experimental studies which have been conducted to date, however, take no account of these contextual factors, often presenting errors for evaluation in isolated sentences. It is perhaps not surprising that these studies have produced conflicting results (see Santos 1987). The appearance of rigour given by the use of descriptive and qualitative statistics may therefore be spurious.