

CLASSIFICATION AND DISTRIBUTION OF LEXICAL ERRORS IN THE WRITTEN WORK OF GERMAN LEARNERS OF ENGLISH¹

RÜDIGER ZIMMERMANN

Marburg University

0. The starting point of our work on lexical errors was dissatisfaction with error correction in universities and high schools: Generally speaking, there is not enough differentiation. We usually do not get much more than *lexical* errors and some *collocational* and *idiomatic* errors; in addition, sometimes indications of *stylistic* appropriateness. Then we may have *word formation* errors. This is the usual kind of correction in written work that most of us will be used to. The main shortcoming of this procedure must be seen in the undifferentiated category of *lexical error*.

Despite recent scholarly contributions towards a differentiation of this category, especially those by Debyser et al. (1967), Levenston/Blum (1977), Arabski (1979) and Ringbom (1978) we continue to find overall ratings. This is the case e.g. in Steinbach's (1981) paper in *PSCL 13*, where almost 50% of all errors (including grammar and spelling) come under the category of *lexical selection*. These lexical selection errors comprise about 80% of all lexical errors (Steinbach 1981: 255).

Similar ratios are typical of everyday practice, but I think that such wide categories are not helpful in any way, neither for learner correction nor for remedial teaching.

Arabski and Ringbom have chosen to tackle the problem using what I would call a mixed procedure, namely combining description and explanation of lexical errors: Their description of approximations often contains hypotheses on the sources. *Transfer, false friends* (Arabski 1979: 32ff; 36 (as "faux amis") or *language switch* and *attempted anglicification* (Ringbom 1978: 89) bear witness to this approach.

1. My approach is similar to the mentioned ones in that I also attempt to

¹ Paper read at the Societas Linguisticae Europaea 16th Annual Meeting, Poznań 1983.

differentiate the concept of lexical error in terms of categories from linguistic semantics. My first question is, then: How relevant are usual semantic categories for the description of one important kind of *real* semantic deviation, namely in learners' errors? Can they account for lexical errors? Which are too wide or too loose? — I think these are by no means trivial questions, since linguistic semantics has often been satisfied with relatively unproblematic examples, e.g. from the field of concrete perceptible objects, whereas social and cultural concepts of higher complexity play a bigger role in foreign language teaching, especially on the advanced level. So my first aim is the exploitation of *one* paradigm. I am making the analytic distinction between linguistic *description* and psychological *explanation* (at least for the time being, in the first phase). In doing so, I hope to do better justice to the *subject matter* of learning by disregarding the learning *process*. In a nutshell: I am interested in a description of the meaning difference between the interlanguage lexeme chosen and the target language lexeme intended, which should be as exact as possible.

While I aim at the possibility of a detailed description of lexical errors I am not sure whether semantic details are also important for the learning process. It may be the case that typical aspects of semantic deviation are more decisive in this respect.

2. A linguistic typology of lexical errors (or: categories of semantic deviation) At this stage I am disregarding depth and gravity of errors. As to the distinction between lexical and grammatical errors I rely on the usual criteria of semanticity and generalization, i.e. the width of rule applicability.

Let me sketch my typology by way of examples from our data (written work from university and, much less, high school classes)².

I am beginning with the confusion of sense relations

(1) Hyp → *Sup

I was thrilled by the white light and the special *smell*
scent/Duft

(2) Hyp → *Sup

I *went* upstairs
jumped/sprang

(3) Sup → *Hyp

Some of his *claims* have become less important
demands/Forderungen

(4) Sup → *Hyp

A dying *colonel* once said
(commissioned) officer/Offizier

² It should be borne in mind that some illustrative sentences contain errors other than the ones discussed.

(1) through (4) are confusions of supernyms and hyponyms in either direction: **Smell* is more general than *scent*; **claim* is more specific than *demand*. (Such cases need no further comment; other analysts handle them in the same way.)

(5) Hyp → *Sup/Style

... which agreed on *getting rid of* all dialects
eradicate/ausrotten

(5) illustrates what has not been brought out clearly enough in previous work: the necessity of double classifications³. **Get rid* is not exact enough in comparison with *eradicate*; that makes it an erroneous supernym, an underspecification. But at the same time the stylistic inappropriateness is felt in this context; *get rid* is too colloquial.

(6) Cohyp₁ → *Cohyp₂

A final decision to *exterminate* all dialects
eradicate/ausrotten

Interlanguage **exterminate* as well as correct *eradicate* are specific terms of forceful removal (roughly equivalent to *destroy*), but the one does not exclude the other. Here co-hyponyms have been confounded.

(7) Het₁ → *Het₂

Stairs were being *washed*
scrubbed/gescheuert

In (7), on the other hand, the meanings of *wash* and *scrub* are mutually exclusive; both heteronyms are specifications of *cleaning*.⁴

(8) Part → *Whole

The injury had to be *operated*
sutured/sewn/genäht

(9) Result → *Process

Theodor Siebs had made phonetic *recordings*
records/Aufzeichnungen

There are other minor groups of sense relation errors, such as (8), where *suturing* or *sewing* is part of the whole action of **operating*, or (9), where the process of **recording* was referred to instead of the resulting *record*.

(10) Field E.

in a *memorial* by the Foreign Office
memorandum/Denkschrift

Errors such as (10) are best described as field errors: Instead of using *memorandum*, a word from the semantic field of "written public communication", the learner wrote **memorial*, which is from the field of "public buildings".

³ Cf. Legenhausen's work on grammatical errors (1975: 18ff.).

⁴ I am still undecided whether cohyponym and heteronym errors should not be classified with feature errors, esp. since there are only few instances in our data.

(11) ? Field E.

The tongue was *stark* swollen.
thickly/dick

(11) gives the impression of a 'macaroni' sentence (Arabski's *lexical shift*), where a German word is used in an English sentence. But this is a causal hypothesis which cannot do justice to the understanding of the native speaker. So it should rather be interpreted as */sta:k/, an adverb-like form with the possible meaning of *utterly*. From this point of view, a word from the field of intensive expressions would have been replaced by a term of extension.

The error types presented so far were based or can be based on a holistic conception of meaning. They are followed by errors which have to be described through partial aspects of meaning, through meaning components, or, in a more formal way, semantic features. I call them feature errors.

(12) Feature E.

I *slided* and fell
slipped/rutschte aus

This is to me a clear case in point: the confusion of **slide* and *slip* is owing to a small difference in meaning. They are loosely synonymous, but in *slip* a component of accident seems to be typical, whereas *slide* is unmarked in this respect.

(13) Field E. or Feature E.

German attempts to establish a standardized language appear much more apolitical, literary, *disinterested*
detached/distanzierter

(13) is rather a border-line case between field and feature errors. It seems feasible to locate **disinterested* in a subfield of "objective attitude" (like *fair*), and *detached* in a subfield of "lack of dedication". On the other hand, the meaning difference can be seen as a case of partial synonymy, which it would be better to grasp through different features.

(14) ? Feature E.

Homecoming is a *nice* adventure
charming/reizendes

Cases like (14) leave me rather insecure: There is a difference in content between **nice* and *charming*, but it seems to be secondary as compared to the pure expression of positive attitudes of different degree and validity.

(15) Word-formation E.

It has been *normed*
become the rule/normiert

This shows one aspect of this error type. Although **normed* is a formal violation of the linguistic norm, it is not a lexical error in the narrower sense: This IL lexeme does not exist in English, but it is a potential word, and communication is not impaired.

(16) The paper is *yellowy*
yellowed/vergilbt

(16), on the other hand, is a word-formation error which is deficient on the formal and semantic level: *-y*-derivations of colour adjectives are impossible, and the meaning is unclear between *yellowish* and *yellowed*. — This category ends my linguistic classification of lexical errors, as far as deviations of meaning are concerned, but I have to add a few more categories which I think are best dealt with under my general topic.

(17) Collocation E.

red-edged eyes
red-rimmed/rotgerändert

(17) looks like a word-formation error, but what is wrong is the collocation: *red-edged* alone is good, but it does not go with *eye*, and it is only *edge* (instead of *rim*) that cannot occur with *eye*:

(18) Non-interpretable form

... which in one of its *momentos* ends like this
memorandum/Denkschrift

(19) and filled the mouth like a *clops*

lump/Kloß

**Momentos* in (18) looks like an English word, but it does not exist, and it cannot be interpreted. **Clops* instead of *lump* (in the throat) is another instance of such noninterpretable forms.

(20) Idiomatic E.

These negotiations *are cutting a great figure*
play an important role/spielen eine bedeutende Rolle

(21) Harald Wilson, ... the *go-getter*|*busy-body*|*Jack-in-the-pot*|*Jack everywhere*
Jack-of-all-trades|*Hansdampf in allen Gassen*

Idiomatic errors such as (20) and (21) need no comment in that their classification poses no problems. But it is very interesting to see what constructions learners use to replace them.

Further minor categories are redundancy and omission; (22) and (23), respectively.

(22) Redundancy E.

At a *rate* speed of five miles a second
Ø|*Geschwindigkeit*

(23) Omission E.

I had bitten through ————— my tongue
edges/Zungenränder

A last important category are paraphrase errors. Let me just indicate a necessary distinction between formally possible ones, such as (24), and syntactically and semantically deviant ones, like (25).

(24) Paraphrase E.

The *outer parts* of my tongue
edges/*Zungenränder*

(25) The tongue was a muscle *which could not be missed*.

indispensable/unentbehrlich

Stylistic errors, e.g. (26) and (27), are again unproblematic as such, but they often have to be double-classified.

(26) Stylistic E.

The doors of the *taverns* are open
pubs/Kneipen

(27) My dealing with *written stuff*

written things/Schriftlichem

Connotative (or associative) deviances are rare: (28) is a case in point.

(28) Connotative E.

I *jerked* upstairs
jumped/sprang

3. Problems of delimitation and overlapping of categories

I think it is evident that this typology raises several problems of delimitation and overlapping of categories. After all, it is organized along four principles:

- Most types are based on a holistic conception of meaning (sense relation errors).
- Feature errors derive from a componential approach.
- Deviances in style, register and connotation are concerned with additional aspects of overall meaning, beyond the conceptual core.
- Then there is the syntagmatic point of view in collocational and paraphrase errors whilst the other types were defined on a paradigmatic basis.

It is clear that sense relation errors can also be formulated in terms of feature differences. As far as it seems feasible I have so far preferred a description using holistic categories, since I think this is more promising from a psychological and acquisitional point of view.

The main difficulty lies in the distinction between feature errors and those field errors where members of similar subfields are dealt with. This group of errors with relatively small meaning differences between the IL and the TL lexemes is one of the biggest in the corpus.

I am trying to delimit along the following line: Wherever IL and TL form can be attributed to fairly distinct subfields, I classify as field error. In such cases it is usually a hierarchically higher feature that accounts for the deviance.⁵ Feature errors, then, are largely occurrences of almost synonymous lexemes. They have often been treated as 'synonyms' in loose terminology.

⁵ I have to admit that the classification as feature or field error depends largely on whether a semantic feature or meaning component lends itself to description.

Paraphrase errors unite deviances of different content and size: They can be attempts to replace a lexeme or a group of words (like and idiom), they can be deviant in content or style, and of course in their formal make-up. They comprise errors in which the inappropriateness or awkwardness appears to be the main trait of the IL expression.⁶

4. Preliminary results⁷

4.1. Distribution of error types.

A preliminary analysis of about 20% of the errors from our corpus of approximately 2000 items, not including paraphrases, but including word-formation errors, has shown the following tendencies (it seems premature at this stage to speak of results):

- About 20% can be classified as sense relation errors, but this has to be done with the caveat that with comparatively many of them a classification as feature errors may be more exact (esp. for verbs and adjectives).
- About 1/3 can be classified as field (or subfield) errors.
- So far only as little as 10% have been classified as feature errors, that is confusion of near synonyms, owing to our tendency to prefer holistic or field- or scheme-oriented descriptions.
- In about 1/5 of all classified errors, the decisive point was seen in the wrong collocation, a surprisingly high ratio for data mostly from university courses and exams, maybe a Marburg-specific phenomenon.

It should be noted that errors belonging with the descriptively most difficult types, namely field and feature errors, constitute together the biggest group (over 40%).

It can be expected that those errors which proved too difficult to classify so far will mostly be incorporated among the feature errors, or double-classified (<10%).⁸

4.2. Form-class-specific errors

A survey on the basis of slightly more than one thousand errors (complete for all versions of several texts from the corpus) showed, among other things, the following results:

- It is not surprising that noun errors are more frequent than verbal, adject-

⁶ It goes without saying that all classifications have been performed using native speaker reactions.

⁷ Our data are so far restricted to written work, translations, essays, and other kinds of relatively free text production. We are aware of the limited value of this kind of data for an overall description of learner language, but we have chosen them, since these types of exercises are at the same time decisive in tests and examinations, however anachronistic this may seem.

⁸ Let me add at this point that a description which would be optimal from a scientific point of view would certainly lead to an error matrix like the one proposed by Le-genhausen (1975: 42).

tival and adverbial deviances; their relative frequency of approximately 4 to 3 to 2 to 1 corresponds largely to general word frequencies in texts of this kind.

But there are some more interesting form-class-specific findings: cf. table 1.

Table 1: Error frequencies

Error type \ Form class	Number of errors (%)			
	N	V	Adj	Adv
Sense rel. E	25	19	10	8
Field E	17	20	22	18
Word-F. E	25	1	12	—
Coll. E	6	18	18	7
Om. E	1	<1	3	13
Par. E	13	25	14	30

- Typical errors among nouns are word-formation, sense relation and field errors, whereas paraphrases are relatively infrequent.
- On the other hand, word-formation errors are extremely rare among verbs; verbal errors are mostly inadequate paraphrases, field and collocation errors.
- Among adjectivals field and collocation errors predominate; paraphrases are again rarer.
- Adverbs of all kinds (but excluding grammatical forms such as intensifiers) show striking peculiarities; almost 1/3 are paraphrases (with *way*, *manner*, *kind* etc.). Field errors are also not rare, and omissions are surprisingly frequent. Finally, there are practically no word-formation errors with adverbs at all.

4.3. Paraphrases and paraphrase errors.

My preceding remarks have made it apparent that paraphrase errors play an important role in our data. It must be kept in mind that these errors are all from written work where there is time enough for planning the utterance. Paraphrases and, consequently, paraphrase errors are even more frequent in oral communication, not only among foreign language learners, but also with native speakers: We often paraphrase a term that does not come to our mind instantaneously, we have something on the tip of our tongue, and we are helped out by others. Paraphrasing, even if it is not as exact and specific as the optimal intended word, is a normal way of everyday communication. It deserves more attention, and, I think, a more positive attitude among foreign language teachers, as do by the way, "creative" word-formations which are not in the dictionary.⁹

⁹ Cf. Zimmermann, in press and Zimmermann/Schneider, in press.

Paraphrases follow certain patterns reflecting again the grammatical and lexical standard of their producers. We use an idealized schema to classify paraphrases, cf. table 2.

Table 2: Paraphrases

meaning-preserving formally correct	* meaning-preserving formally deviant
* meaning-changing formally correct	** meaning-changing formally deviant

Illustrative examples:

Good: **which had the colour of smoke/smoke-coloured*

Only formally deviant: **in a polite form/politely*

Only meaning-changing: **never greedy/always modest*

Completely deviant: **stood *horizontal/stuck crosswise.*¹⁰

Let me indicate some typical paraphrase structures:

a. Paraphrasing nouns:

$$N \rightarrow \text{Sup} + \left\{ \begin{array}{l} \text{of NP} \\ \text{inf clause} \end{array} \right\}$$

pragmatist → **man of pragmatics*

nurse-maid → **woman to look after the children*

$N \rightarrow \text{Adj} + \text{Sup}$

pragmatist → **pragmatic man*

$N\text{-compound} \rightarrow N + \text{rel clause}$

post-war (hopes) → **(hopes) coming up after the war, *(hopes) people had for the time after the war*

b. Paraphrasing verbs:

$V_x \rightarrow V_{\text{link}} + N_x$

Imagined → **had the imagination of*

$V_x \rightarrow V_Y + \text{PP}$

flipping (across the surface) → **appearing in a quick movement ...*

c. Paraphrasing adjectives

$\text{Adj} \rightarrow \text{of} + \text{NP}$

social-democratic → **of the social democrats*

$\text{Adj} \rightarrow \text{rel clause}$

icy → **which had become ice*

$\text{pol adj}_1 \rightarrow \text{neg} + \text{pol adj}_2$

bent → **not upright*

d. Paraphrasing adverbs

Only by prepositional phrases

permanently → **without interruption*

¹⁰ Non-interpretable paraphrases are grouped with the meaning-changing ones, redundant ones go with redundancy errors.

All these paraphrases were marked as incorrect or somehow deviant, but I think it is clear that many of them would go unnoticed in even educated oral communication. They should be encouraged rather than penalized, and they ought to be furthered by "paraphrasing exercises", which, to my knowledge, do not exist in our universities.

Let me conclude these remarks on paraphrases by pointing out their particular relevance for the expression of idioms, where good paraphrases can of course not be the optimal solution, but nevertheless "idiomatic English". Cf. **master of all situations* or even **man capable of managing everything to *whirling-pool in all lanes* **James everywhere* or even **busy Jack in all trades*: the first two are certainly much more acceptable than the last three (cf. example 21).

5. Further perspectives

We are now also investigating the relation between communicative damage and sheer irritation in native speakers. Then we are beginning to complement our error files by causal hypotheses, including associative sources in L_1 and L_2 (Zimmermann, to appear).

This should eventually lead to partial analyses of selected errors in terms of learner language. We have some fine examples of only seemingly correct forms which are in reality covert lexical errors: The fraction of correct *extrapolated* for German *extrapoliert* does not look so convincing anymore in the light of multiple incorrect forms such as **excluded*, **extracted*, **exposed*, **extraposed* and even **extrapolished*.

REFERENCES

- Arabski, J. 1979. *Errors as indicators of the development of interlanguage*. Katowice: Uniwersytet Śląski.
- Blum, S. and Levenston, E. 1979. "Lexical simplification in second language acquisition". *Studies in Second Language Acquisition* 2. 43 - 63.
- Corder, S. P. and Roulet, B. (eds). 1977. *The notion of simplification, interlanguages and pidgins and their relation to second language pedagogy*. Geneva: Droz.
- Debyser, F., Houis, B. and Rojas, C. 1967. *Grille de classement typologique des fautes*. Paris: B. E. L. C.
- Legenhausen, L. 1975. *Fehleranalyse und Fehlerbewertung*. Untersuchungen an englischen Reifeprüfungsnacherzählungen. Berlin: Cornelsen.
- Levenston, D. and Blum, S. 1977. "Aspects of lexical simplification in the speech and writing of advanced adult learners". In Corder, S. P. and Roulet, B. (eds). 1977. 51 - 71.
- Ringbom, H. 1978. "The influence of the mother-tongue on the translation of lexical items". *ISB* 3. 80 - 101.
- Steinbach, H. R. 1981. "On the classification of errors in translation papers with some consideration of interference phenomena". *PSiCL* 13. 249 - 59.
- Zimmermann, R. In preparation. "Content-oriented and form-oriented lexical errors", to appear in *IRAL*.