

Vocabulary Deviations during Second Language Lexical Processing

Eda Duruk

English instructor, Ph.D.,
School of Foreign Languages,
Pamukkale University,
Denizli, Turkey

Turan Paker

Assoc. Prof. Dr.,
English Language Teaching Department,
Pamukkale University,
Denizli, Turkey

Abstract

The present study is a descriptive, quantitative study which measures second language (L2) learners' extent of lexical errors in variety contexts. In the study a varied corpus was gathered by means of participants' (n=30) written papers. Turkish-speaking learners of English from the pre-intermediate level (n=100) completed a productive and receptive task that presented contrasting forms of various lexis obtained by the corpus. The results of both tasks indicate that learners are often unable to detect the lexical errors and correct them. In particular, learners have significant difficulty in word formation, word choice, collocations, prepositions, and part of speech. These difficulties are interpreted as stemming from similarities as well as generalizations. The implications of the findings are discussed accordingly.

Keywords: interlanguage, vocabulary acquisition, error analysis, second language acquisition

Introduction

Language learning begins with vocabulary. Learners acquire them as the first linguistic items (Llach, 2005). Vocabulary is the key component of communication. Words are the tools to express meanings and without them, grammar is just a meaningless abstract construct of rules (Llach, 2005).

Second language (L2) vocabulary acquisition has drawn lots of attention in the increasing number of studies in second language acquisition (SLA). Learners of L2 usually know the extent to which deficiencies in their vocabulary knowledge affect their communication skills. They need lexical items to give the message they wish to be comprehended (Kim, 2008). Consequently, L2 learners sometimes face difficulties while trying to choose the right lexical items. Therefore, investigating the ways in which words are deviated is of special interest. Such kind of investigation helps remedy those deviations; thus, makes the intended message more understandable.

What does it mean to know a word?

According to Bogaards (2001), the notion of “word” is unclear. It is not always straight forward to ignore totally and to have partial or full knowledge of a word. Words carry many shapes and they can be grouped in many ways. They reflect various functions and fulfill distinct roles.

Knowing a word is more than just learning its meaning and form (Schmitt & Meara, 1997). Following Richards (1976, cited in Schmitt & Meara, 1997) who made the first attempt to explain the types of knowledge necessary to fully know a word, Nation (1990, p.31, cited in Schmitt & Meara, 1997) improved the list in the following way:

1. The spoken form of a word.
2. The written form of a word.
3. The grammatical behavior of the word.
4. The collocational behavior of the word.
5. How frequent the word is.
6. The stylistic register constraints of a word.
7. The conceptual meaning of a word.
8. The associations a word has with other related words.

When all these types of knowledge are acquired, a learner can hopefully use the word like a native-speaker in various contexts. Schmitt and Meara (1997) also noted that grammatical knowledge carried by a word can give lots of clues, from its morphological features to its word class. Widdowson (1978, cited in Liu & Shaw, 2001) mentioned that both use (pragmatic function) and usage (form) are required in learning a language. The key to successful native-like performance lies in the mastery of lexical relations – collocations, lexical phrases, fixed phrases (Liu & Shaw, 2001).

Literature Review

Concurrently, studies on vocabulary and its relationship to the writing process have come to the fore. However, as Engber (1995, cited in Llach, 2005) points out, not most of these studies are directly related to lexical errors and their impact on ESL writing tasks. Lexical errors and lack of lexical knowledge are believed to have a great influence on communication. Akande et al. (2006) also stated that concerning the mastery of English lexis, there are some constraints which ESL learners are faced with, and they are believed to affect learners communicative competence. Read (2004, p.146, cited in Akande et al., 2006) supported this by his following statement:

Second language learners are typically conscious of the extent to which limitations in their vocabulary knowledge hamper their ability to communicate effectively in the target language, since lexical items carry the basic information load of the meanings they wish to comprehend and express.

In the present study, lexical errors distort written communication, which obstruct the meaning of the intended message. It is of great importance to determine what a lexical error is. According to Duskova (1969) this can be unexpectedly easy at times, since the degree of deviation from the normal form is such as not to leave any doubt of the unacceptability of the form in question. On the other hand, when forms displaying a low degree of deviation are concerned, it can be difficult to decide whether to regard them as errors. The following examples of Duskova (1969, P.27) exemplify the distinction in a clear way:

1. I *obtained* many fruitful remarks (received)
2. May I *ask you for a kindness* (ask a favor of you)
3. I *write* with two workers from this Institute (correspond)
4. The solution proposed in this article should not be *concerned*

as final (considered)

They are all samples of lexical errors. Sentences 1 and 2 are still understandable, as Duskova (1969) states; on the other hand, sentences 3 and 4 obstruct the meaning of the message.

Bogaards (2001) drew attention to the content words and underlined that such words can also be problematic as they carry lots of different meanings. He made distinction by the following examples:

1. Our neighbors are throwing a party tonight.
2. They were very grateful to the rescue party.
3. The Conservative Party has lost many votes.
4. The lawyer refuted the arguments of the other party.
5. Your party is on the line.

According to Bogaards (2001), the sense of the each word given above is quite different even though some of them can be viewed as having the same root diachronically for ESL learners. Thus, the content words like “party” can have extremely rich meaning potential, and so they are open to misuse.

Lots of studies have been conducted on classifying lexical errors. Corder (1973, cited in Llach, 2005) suggested that a word has mainly four properties; semantic, syntactic, phonological and pragmatic. According to Llach (2005), lexical errors are due to the wrong word choice, the use of a L1 word, and confusion of similar words.

Akandeet al. (2006) conducted a study in which they investigated the lexical errors in the English of some Technical College students by applying a 20-item fill-in Multiple Choice test and an essay writing exercise. They noted lexical errors of various types. They concluded that the most common errors were generalization errors, wrong lexical choice, similarity errors, wrong word formation, collocation errors and duplication errors, respectively.

Zyzik and Azevedo (2008) examined ESL learners’ knowledge of word class distinctions in a variety of syntactic contexts. The participants completed a receptive task which covered contrasting forms belonging to the same word family. The results showed that participants were often unable to distinguish among word classes. Indeed, the most difficulty was seen in discriminating between adjectives and nouns.

Methodology

The Research questions

Learners of second language face a daunting situation: learning multi-aspects of lexis and using them appropriately in a given context. Learners from different proficiency levels make different lexical errors.

This study investigates lexical errors and their types depending on a varied corpus formed by Turkish-speaking learners of English from the pre-intermediate level. The study aims (1) to detect and classify the lexical errors by means of receptive and productive tasks, (2) to find the most common errors by examining the frequency of each error, and (3) to account for the possible sources of the problems and suggest solutions to remedy them.

The research questions are formulated in the following way:

- 1) How many types of lexical errors do pre-intermediate students make?
- 2) What are the common lexical errors pre-intermediate students make? Which errors are the most common ones?
- 3) Are lexical errors limited to production, or do L2 learners also face it in the receptive mode?

With respect to the third research question, it is hypothesized that learners will make errors in both productive and receptive tasks. If this hypothesis is confirmed, then it would indicate that lexical errors are not merely limited to performance. The second hypothesis is that the frequencies of each error type detected in the productive task will be higher than those of detected in the receptive task.

Participants

The study was conducted in a western university in Turkey. 130 students participated in the study. 30 of them were used for obtaining the corpus and the remaining 100 participated in the main study. They were all preparatory students enrolled in the same courses offered for pre-intermediate level. Their L1 was Turkish, and none of them had been exposed to English in a foreign country. Their ages ranged from 19 to 22. The length of learning English as a second language varied from 3 to 7 years. At the beginning of the first semester, they were placed in the pre-intermediate level by applying a placement test. There were mainly four parts in the test; listening, language use, reading and writing. The first three parts of the exam included multiple choice questions in which there were 4 distracters apart from the correct choice. In the listening part, there were 10 questions which were totally 20 points. Language use part included 30 questions and each of them was one point. The reading part had 15 questions which were totally 30 points. As for the writing part which was 20 points, the participants were given a picture of a crowded street and asked to describe the picture in detail. The students who scored between 50 and 60 points were placed in pre-intermediate level. The courses they had taken so far at the university were core, speaking and listening, reading and writing. The writing course was a part of the curriculum which covered four hours a week. The study specifically examined lexical errors in participants' process paragraph writing. Two weeks were devoted to process paragraph writing. The section was taught by a lecturer who was not included in data analysis.

Instruments

A preliminary step in developing a reliable instrument was to gather lexical errors produced by L2 learners. Given that Whitley (2004) documented many such errors in in-class writing papers, a similar process was followed in the study in which learners' process paragraphs were examined. Learners' (n=30) quiz and midterm papers of the same single semester were gathered (see Appendix A&B). Learners were given 40 min to write on the assigned topics during two different class periods. The second paragraph was asked to be written 3 days later than the first one. They had no recourse to dictionaries or any other materials while performing the written production tasks. Given that open-ended nature of the written tasks, learners could certainly avoid what they perceived to be difficult L2 constructions or unfamiliar lexical items. Nonetheless, the researcher deemed two written responses from each learner to be sufficient for gathering a varied corpus of lexical errors. For the sake of inter-scorer reliability, these two corpora were analyzed by two lecturers (the researcher as one of the lecturers). As a standard reference, Longman, *The Dictionary of Contemporary English* was used. Both scorers used the list of criteria for coding the written data (see Appendix C). They exhibited a 92% agreement rate in error identification. Of the total errors identified, there was 100% inter-scorer agreement on the type of error (wrong word formation, wrong word choice, prepositional mistake, etc.). While forming the corpus, lexical errors were categorized in terms of type of error and the frequency list of each error type was formed to see how often each different lexical error occurred in the written works of the participants. The data were recorded via the SPSS 16.0 program. Representative samples from the learner-generated corpus are provided in Appendix D. The written production tasks revealed that certain error types were much more prevalent than others (they were further discussed in the results and discussion part). After coding the written production data, the researcher designed a productive and a receptive task based on the four most common error types. Both of the tasks included target items which were taken directly from learners' written production. The sentences of both tasks were also taken from the corpus. A few modifications were made where the sentences did not provide the necessary context. The use of real data in the elaboration of the tasks was important because it allowed

the researcher to work with lexical items and constructions familiar to learners rather than contrived sentences generated for research purposes.

The productive task, accuracy judgment task (AJT), consisted of 15 sentences (see Appendix E). The learners were asked to determine the accuracy of 15 sentences and to correct the inaccurate sentences. Of all the sentences, 12 were inaccurate. The remaining 3 sentences were accurate and they functioned as distracters (sentences 4, 8 and 11). Each of the four types of errors was randomly distributed to 3 sentences. The items and the order of the correct choices in the AJT were randomized. Sentences 1, 10 and 13 test errors on part of speech; sentences 2, 7 and 14 test prepositional errors; sentences 3, 12 and 15 test wrong word formation; and sentences 5, 6 and 9 test wrong word choice.

The receptive task, a forced choice task (FCT), included 12 sentences and 12 fillers along with three choices (see Appendix F). The learners were asked to circle the correct choice that could complete each sentence. If the learner felt that neither form was appropriate, the learner could circle *neither*. The advantage of this format over a simple binary one is that it reduces the likelihood of guessing. Similar to the AJT, each of the four types of errors was randomly distributed to 3 sentences. Sentences 1, 8 and 9 test wrong word formation; sentences 2, 10 and 11 test prepositional errors; sentences 3, 5 and 6 test errors on part of speech; and sentences 4, 7 and 12 test wrong word choice. To ensure the reliability of the instrument, a measure of internal consistency was computed. The 12 target items yielded an alpha-reliability coefficient of .79, indicating a good level of interrelatedness among the items.

Results

Corpus

As discussed in the development of the instrument, while forming the corpus, lexical errors were categorized in terms of type of error and the frequency list of each error type was formed by the two scorers to see how often each different lexical error occurred in the written works of the participants. The first set of statistical analyses was carried out on the categorization of participants' different error types. The next step was to consider the overall frequency and percentage of each error type. The results are shown in Table 1.

Table 1. Frequency and percentage statistics for total error occurrence of each error type

Error Type:	TOTAL	Collocation	Preposition	Part of Speech	Wrong Word Formation	Wrong Word Choice	Grammatical Error
n	277	22	55	49	54	62	35
%	100%	7.8%	20%	18%	19.4%	22.3%	12.5%

As can be seen in Table 1, lexical errors frequently detected in the process paragraphs of the participants were on collocations (7.8%), grammar (12.5%), part of speech (18%), word formation (19.4%), prepositions (20%) and word choice (22.3%).

The Accuracy Judgment Task (AJT)

There were 15 sentences in the task. For each of the 3 correct sentences, participants received 1 point for putting a tick and 0 for putting a cross. For the rest 12 sentences, participants received 2 points for detecting and correcting the error; 1 point for detecting but not being able to correct the error; and 0 for not even detecting the error of each sentence. The second set of statistical analyses was carried out on each participant's total score on the AJT. The results are shown in Table 2 along with the mean score. The results

for each of the four conditions were further examined. There were 3 items for each of the 4 conditions. Frequencies and percentages for each four condition are listed in Table 2.

It can be understood from Table 2 that out of 27, the mean score of the participants (n= 80) is 6, 24. Participants were unable to either detect and/or correct most of the lexical errors in the given sentences.

Table 2. Frequency and percentage statistics for each condition on AJT

Type of Error:	Score:									
	0		1		2		3		4	
	n	%	n	%	N	%	n	%	n	%
Part of Speech	33	41	15	19	20	25	10	13	2	3
Wrong Word Formation	16	20	26	33	16	20	10	13	10	13
Wrong Word Choice	41	51	6	8	16	20	10	13	7	9
Prepositional Error	35	44	4	5	32	40	5	6	4	5
The mean score: 6,24										

It can be inferred from the Table 2 that 33 participants could not spot the errors on part of speech. 15 of them only detected one of the errors; thus, they scored 1 point. 20 of them scored 2 points, 10 scored 3 points. 2 of them scored 4 points. As for word formation, 16 participants could not spot the errors. 26 of them only detected one of the errors; thus, they scored 1 point. 16 scored 2 points, 10 scored 3 points. Other 10 scored 4 points, and 2 scored 5 points. With respect to word choice, 41 participants could not spot the errors. 6 of them only detected one of the errors; thus, they scored 1 point. 16 scored 2 points, 10 scored 3 points and 7 scored 4 points. 35 participants could not spot prepositional errors. 4 of them only detected one of the errors; thus, they scored 1 point. 32 scored 2 points, 5 scored 3 points and 4 scored 4 points.

The mean scores of errors on part of speech, wrong word formation, wrong word choice and prepositional errors are 1,16, 1,73, 1,20 and 1,24, respectively.

Table 3. Frequency and percentage statistics for correct sentences on AJT

	Score:	n	%
	Correct sentences	0	32
1		27	34
2		17	21
3		4	5
The mean score: 91			

It can be seen from Table 3 that 32 participants could not spot the correct sentences. 27 detected 1 of them; whereas 17 detected 2 of them. 4 identified all of the correct sentences. The mean score of the correct sentences is , 91.

The Forced-Choice Task (FCT)

On each of the target items (n=12), participants received 1 point for choosing the correct choice and 0 for any of the remaining possible choices. The third set of statistical analyses was carried out on each participant's total score on the FCT. The results for each of the four conditions are shown in Table 4 along with the mean score. There were 3 items for each of the 4 conditions, for a total of 12 target items. Frequencies and percentages for each four condition are listed in Table 4.

Table 4. Frequency and percentage statistics for each condition on FCT

Type of Error:	Score:							
	0		1		2		3	
	n	%	n	%	N	%	n	%
Part of Speech	6	8	44	55	27	34	3	4
Wrong Word Formation	10	13	32	40	28	35	10	13
Wrong Word Choice	2	3	31	39	34	43	13	16
Prepositional Error	13	16	29	36	30	38	8	10
The meanscore: 5,95								

It can be understood from Table 4 that out of 12, the mean score of the participants (n= 80) is 5, 95. Most of the participants identified almost half of the correct choices. It can be seen from the Table 4 that 6 participants could not choose the correct choice on part of speech. 44 of them answered only one of the 3 questions correctly; thus, they scored 1 point. 27 of them answered 2 questions correctly; thus, they scored 2 points. 3 answered all 3 questions correctly; thus, they scored 3 points. As for word formation, 10 participants could not choose the correct choice. 32 of them answered only one of the 3 questions correctly; thus, they scored 1 point. 28 of them answered 2 questions correctly; thus, they scored 2 points. 10 answered all 3 questions correctly; thus, they scored 3 points. With respect to word choice, 2 participants could not choose the correct choice. 31 of them answered only one of the 3 questions correctly; thus, they scored 1 point. 34 of them answered 2 questions correctly; thus, they scored 2 points. 13 answered all 3 questions correctly; thus, they scored 3 points. 13 participants could not choose the correct choice on prepositions. 29 of them answered only one of the 3 questions correctly; thus, they scored 1 point. 30 of them answered 2 questions correctly; thus, they scored 2 points. 8 answered all 3 questions correctly; thus, they scored 3 points.

The mean scores of errors on part of speech, wrong word formation, wrong word choice and prepositional errors are 1,34, 1,48, 1,72 and 1,41, respectively.

Discussion

With respect to the first and second research questions, the writing papers of the participants revealed that the following 6 types of lexical errors were made; (1) collocation error, (2) prepositional error, (3) errors on part of speech, (4) wrong word formation, (5) wrong word choice, and (6) grammatical error. These findings are parallel with the findings of Akande et al. (2006). They noted 6 types of lexical errors; (1) collocation error, (2) duplication error, (3) generalization error, (4) similarity error, (5) wrong lexical choice, and (6) wrong word formation. Similarly, some of the errors also stemmed from similarity and generalization. Some of the error samples stemming from similarity are discussed as examples. The first sample error in the sentence “*Anxiety is related to the physiological effects*” (instead of psychological) is due to spelling similarity. Moreover, the wrong usage of “do” in the sentence “*I will do a party*” (instead of have) is due to the fact that its L1 counterpart is similar in terms of meaning. The reason of such kind of errors is lexical gap, in which the meaning of one lexical item in L1 has two or more counterparts in L2. Llach (2005) also described such kind of errors as the confusion of similar words between L1 and L2. On the other hand, Whitley (2004) defines these errors as transfer errors in which usages are apparently carried over from learners’ L1. As for generalization errors, the following sentence from the corpus can be given as an example: “*The teacher tached the subject effectively*” (instead of taught). According to acquisition hierarchy, the acquisition of some structures precedes the other ones. Regular past tense marker “-ed” are

among the ones acquired first. Thus, when irregular verbs such as “teach” are faced, L2 learners naturally mark it with the regular past tense marker. All of the other types of lexical errors such as collocation error, prepositional error, wrong word formation, etc. can be regarded as interlanguage errors.

In response to the second research question, it was observed that, given the process paragraph writing, learners were least accurate in word choice, prepositions, word formation and part of speech. Their frequencies were listed as 62, 55, 54 and 49, respectively. The finding that the most common error is the wrong word choice correlates with what Akandeet al. (2006) suggested. They also placed wrong lexical choice in the second most common errors. The finding that L2 learners have more difficulty in word choice, prepositions, word formation and part of speech can be explained by their proficiency levels. The participants were selected from pre-intermediate level. Learners from this level generally start to take risks and try to produce something new, most of which results in interlanguage errors. Whitley (2004) also indicated that lexical errors are not solely limited to the need of producing a word that corresponds to a given meaning, but also stemming from the need to recast that meaning in a different syntactic category, e.g. the noun or adverb equivalent of a given verb.

As for the last research question, the mentioned frequencies for each condition on AJT and FCT indicate that learners also experience some problems with lexis in the receptive mode. However, they are much less than those found in the productive mode. These results from both tasks confirm the first hypothesis. With respect to the second hypothesis, when the relationship between the total scores for each condition on AJT and FCT and participants’ frequencies on these conditions along with the mean scores is considered, it is clear that learners were least accurate while dealing with the productive task. Therefore, the second hypothesis is also confirmed. Zyzik and Azevedo (2008) mentioned that word class errors are not merely a performance phenomenon but are reflective of competence deficit. These findings are similar with the present study. However, in contrast to the present study, Whitley (2004) suggested that errors on part of speech are simply a performance problem.

Conclusion

This study provides insight into the nature of L2 learners’ difficulties with lexical items. The results of both tasks indicate that learners commit lexical errors on both productive and receptive tasks. In particular, learners have significant difficulty in word formation, word choice, prepositions, and part of speech. These difficulties are interpreted as stemming from similarities as well as generalizations. Interlanguage errors were among the most common ones. However, as Chan (2004) indicates, it is probable that avoidance behavior is a factor that affects students’ performance in such a way that they tend to avoid using certain lexical items for fear of making errors.

Given that L1 interference can be the source of errors, especially in the earlier phase of their interlanguage, the adverse effects of negative transfer should be considered. Both researchers and practitioners can contribute to helping learners overcome these effects. Because errors are indicative of the learners’ interlanguage, representing systematic deviations from the norms of the target language in the language learning process, SLA researchers can conduct corpus-based empirical research (such as the present study) with a view to establishing a comprehensive taxonomy of lexical errors for teachers’ reference. If similar error types are classified in the same category, learners may be more able to generalize the nature or the causes of their errors and teachers can be able to provide more effective corrective feedback. At the same time, “a taxonomy can also help to guide syllabus designers and curriculum writers to anticipate learning problems, focus their remedial efforts and to organize error correction activities” (Li & Chan, 1999, p. 81).

Limitations and Future Research

The first limitation concerns the number of questions applied for each error type during the AJT and FCT. Only three sentences for each error type were used in each task. Thus, the results can be only indicative of the possible error types. The setting was limited to a western university in Turkey in order to promote a more homogeneous sample and limit the influence of other extraneous variables. Therefore, more study is required to replicate the findings of the study. The participants were from pre-intermediate level. Thus, further study can be conducted with other proficiency levels. The data collection was based on the written compositions of the participants, not on their oral work. Finally, mistakes and errors committed by the participants were all counted as errors in the study.

References:

- Akande, A. T., Adedeji, E. O. & Okanlawon, B. O. (2006). Lexical Errors in the English of Technical College Students in Osun State of Nigeria. *Nordic Journal of African Studies*, 15 (1). 71-89.
- Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *SSLA*, 23. 321-343.
- Duskova, L. (1969). On sources of errors in foreign language learning. *IRAL*, 7 (1). 11-36.
- Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 58:2, 285-325.
- Li, D. C. S., & Chan, A. Y. W. (1999). Helping teachers correct structural and lexical English errors. *HongKong Journal of Applied Linguistics*, 4 (1). 79-102.
- Liu, E. T. K. & Shaw, P. M. (2001). Investigating learner vocabulary: A possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *IRAL*, 39. 171-194.
- Llach, M. P. A. (2005). The Relationship of Lexical Error and their Types to the Quality of ESL Compositions: an Empirical Study. *PortaLinguarum*, 3. 45-57.
- Schmitt, N. & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *SSLA*, 20. 17-36.
- Whitley, S. (2004). Lexical errors and the acquisition of derivational morphology in Spanish. *Hispania*, 87. 163-172.
- Zyzik, E. & Azevedo, C. (2008). Word class distinctions in second language acquisition. *SSLA*, 31. 1-29.