A Learner Corpus-Based Study on Verb Errors of Turkish EFL Learners

Cem Can

Correspondence: Cem Can, Cukurova University, Turkey.

Abstract

As learner corpora have presently become readily accessible, it is practicable to examine interlanguage errors and carry out error analysis (EA) on learner-generated texts. The data available in a learner corpus enable researchers to investigate authentic learner errors and their respective frequencies in terms of types and tokens as well as contexts in which they regularly occur. The need to consider these authentic learner errors in the design of useful language learning programs and remedial teaching materials has been widely emphasized by many researchers (see e.g., Juozulynas, 1994; Mitton, 1996; Cowan, Choi, & Kim, 2003; Ndiaye & Vandeventer Faltin, 2003; Allerton et al., 2004). This study aims at analyzing inflectional, derivational and word form errors for verbs produced by Turkish EFL learners across six distinct proficiency levels, A1-A2; B1-B2; C1-C2, as defined by Common European Framework of Reference for Languages (henceforth CEFR) (Council of Europe, 2001). The corpus used in this study is the Cambridge Learner Corpus (CLC), the largest annotated test performance corpora which enables the investigation of the linguistic and rhetorical features of the learner performances in the above stated proficiency bands. The findings from this study seem to indicate that, across different proficiency levels and across different registers and genres, the most common verb error categories are incorrect tense of verb (TV), wrong verb choice (RV), wrong verb form (FV), missing verb (MV), and verb agreement (AGV) errors. This study's approach uses the techniques of computer corpus linguistics and has its roots in the Error Analysis framework as proposed by Corder (1971): identification, description, classification and explanation of errors.

Keywords: learner corpora, verb errors, corpus-based error analysis, Turkish EFL learners

1. Introduction

Errors are essential parts of human learning processes just as error correction is an indispensable part of the teaching process and, unlike mistakes, systematic. They characterise our linguistic competence (Corder, 1967) and may reveal clues about the learners' learning strategies. Error detection is the process of determining whether the semantic content and the linguistic form of learners' communicative performances are erroneous or deviant from the norms. This process can be executed by examining anecdotal evidence or by classroom observation (Swan & Smith 2001; Kızıl & Kilimci, 2014b). However, as it encompasses a large number of examples of comprehensive range of error types, a learner corpus is a more dependable option to investigate the (relatively) naturally occurring learner errors. Compiled in a systematic way, it enables the researchers to target error typologies specific for learners from various mother tongue backgrounds. Having a specific focus, a specialized corpus could be utilised to advise language teaching. For example, "descriptions of structure, reliable models of usage, how words and phrases are translated," determining the essentials in a syllabus, and identifying learners' errors can all be developed and/or supported by focusing on specific genres and subgenres (Sinclair, 2001, p. xiii).

In spite of the diverse nature of their methodologies, theoretical assumptions, data and terminology, Contrastive Analysis (CA), Error Analysis (EA) and Interlanguage (IL) studies in Second Language Acquisition can be recognised as setting an analogous pedagogical goal, which is the scientific study of learner language in order to achieve a better understanding of language learning processes. The most evident distinction among them is the approach to learner's performance and the errors. While studies in CA intrinsically do not take any position on this issue, traditional EA considers errors to be detrimental and tries to eliminate them; in the framework of IL, the deviations from the TL norms are considered as an essential component and indicator of the learner's system. Computer-aided error analysis, which is based on learner corpus, on the other hand, "represents a major improvement over traditional EA, not least because instead of considering errors in isolation, it sees them as part of a system which includes both correct and incorrect uses" (Gilquin and Granger, 2015, p.427), combining the underlying principles of IL and EA. Regarding the use of learner

corpus based error analysis and the weaknesses of traditional error analysis, Ellis (1994) emphasises "the importance of collecting well-defined samples of learner language so that clear statements can be made regarding what kinds of errors the learners produce and under what conditions" and states that "many EA studies have not paid enough attention to these factors, with the result that they are difficult to interpret and almost impossible to replicate" (p. 49). Dagneaux et al., (1998) also criticises the nature of the data and related error typologies used by traditional error analysis and lists five limitations in this regard:

- Limitation 1: EA is based on heterogeneous learner data;
- Limitation 2: EA categories are fuzzy;
- Limitation 3: EA cannot cater for phenomena such as avoidance;
- Limitation 4: EA is restricted to what the learner cannot do;
- Limitation 5: EA gives a static picture of L2 learning (p. 164).

Schachter (1974) also points out that, EA can only be used to identify and analyse errors that learners actually produce, and ignores potential errors that learners do not make because they avoid difficult structures. Due to these limitations of traditional EA, learner-corpus-based approach to error analysis has been considered as a new direction in the field with its pedagogical implications being perceptible as a useful methodology to guide Foreign Language (FL) teaching.

Studies focusing on the variation of errors along with language proficiency reveal contradictory results. The general assumption would be that the learners to make fewer errors as their proficiency level increases and the errors are related to proficiency (Engber, 1995; Grant & Ginther, 2000; Kızıl & Kilimci, 2014a; Olsen, 1999). However, some studies reveal that errors are rather common even at the advanced level (e.g. Dagneaux et al. 1998). Moreover, it is usually expected that learners to make different types of errors as they get further. However, in her longitudinal study, Chamot (1987) reveals a great deal of fluctuation in the error typology across the levels and persistence related to some errors. Some studies also stated that the further proficiency level the learners attain, the lower the L1 influence would be (Taylor, 1975). On the other hand, some other research studies claim that that some errors can only appear at a higher level of proficiency (Kellerman, 1983).

In their computer-driven error analysis study of the French EFL learners' errors across levels in International Corpus of Learner English (ICLE), Dagneaux et al., (1998) analysed verb errors. The verb error categories they focused included auxiliary errors, tense errors, finite/non-finite errors, concord errors, morphology errors, and voice errors. They revealed that the auxiliaries are most prominent error category in the intermediate group but is superseded by the tense category in the advanced group. They concluded that the progress in this regard is much more marked for auxiliaries than for tenses.

Verbs, often conveying the main idea of a sentence (Schuler, 2005), are the fundamental element for describing events and expressing relations between entities (Chklovski & Pantel, 2004). The verb is the core of the clause and has a relational meaning, relating to one or more or arguments to an event. In general, verbs are more complex than nouns and tend to represent a greater cognitive load on processing than nouns. Another basic characteristic of prototypical verbs is that they crucially involve change through time, whereas concrete nouns tend to be stable across time. As verbs, to a greater extent, are language-specific and demonstrate variation with respect to meaning across languages (Gentner & Boroditsky, 2001), they represent a major problem for foreign language learners.

For this reason, focusing on the verb errors, this corpus-based study aims at analyzing inflectional, derivational and word form errors for verbs produced by Turkish EFL learners across six distinct proficiency levels, A1-A2; B1-B2; C1-C2, as defined by henceforth CEFR.

2. Methodology

Bardovi-Harlig and Bofman (1989) states that many of the studies of grammatical development in second language acquisition could be classified into two:

(a) studies that examine formal features and

(b) studies that seek to gauge overall progress by a developmental index. (P. 18)

The current study falls into the second category. It aims at detecting and analysing the interlanguage characteristics of developmental verb errors in Turkish EFL learners using a substantial body of spoken and written data. As Biber, Conrad, and Reppen (1998) emphasise, the studies using large data sets may lead to significant improvements in the understanding of language acquisition. For this reason, analysing Cambridge Learner Corpus (CLC), probably the largest annotated learner corpus (29,266,800 words), this study aims to contribute to the emerging body of research in learner-corpus-based error analysis and foreign language learning with its investigation of the verb errors of Turkish

learners of English as a Foreign Language (EFL). The verb error types examined include inflectional, derivational and word form errors for verbs produced by Turkish EFL learners across six distinct proficiency levels, A1-A2; B1-B2; C1-C2, as defined by CEFR. The corpus used in this study is the Turkish EFL learner subcorpus of CLC. CLC consists of texts produced by learners from 130 different mother tongue backgrounds as part of the University of Cambridge English for Speakers of Other Languages (ESOL) examinations. All the learner errors are annotated and each text is provided with the metadata such as the mother tongue background, nationality, age, gender, CEFR level of the learner, format and the register of the text, and the test's name.

The Turkish EFL learner subcorpus extracted from CLC consists of 801.940 words. The breakdown of the corpus is as follows.

Table 1. The breakdown of Turkish EFL learner subcorpus

CEFR level	Number of words
A1 - A2	289.888
B1 - B2	360.750
C1 - C2	151.302
Total	801.940

According to Granger (2003), an ideal error taxonomy to be used in error analysis should be consistent, informative, flexible and reusable. As stated by (Heift and Schulze), most error taxonomies have the following problems:

- Coverage: some errors may be left unclassified when the taxonomy is applied to more texts or future analysis, due to the impossibility of error anticipation;

- Homogeneity: it is difficult to achieve a comprehensive and meaningful description of errors; and,

- Classification characteristics: many categories are language dependent or cannot be easily distinguished

In this study, the verb error typology used is based on the one developed by CLC. Learner errors in CLC are tagged using the following convention (CUP, 2014, p. 14-15):

<#CODE>wrong word|corrected word</#CODE>

The verb error codes are based on a two-letter coding system (except the agreement of verb error - AGV) in which the first letter represents the type of error, while the second letter identifies the word class, which is Verb (V). Verb errors tagged using this convention and the examples of the related errors made by Turkish learners are as follow:

- AGV: Agreement of verb error: the verb does not agree with the subject

There **<#AGV>is | are </#AGV>** so many shops near my house. (Doc ID: 760382; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: A2 Waystage; Year: 2000; Age: 13; Gender: F)

The food **<#AGV>are | is </#AGV>** very important. (Doc ID: 562340; Nationality: Turkey; Exam: KET; CEFR Student Level Performance: B1 Threshold; Year: 1997; Age: 20-30; Gender: F)

Finally the exhibition and the museum **<#AGV>was | were </#AGV>** opened. (Doc ID: 565528; Nationality: Turkey; Exam: CPE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 2000; Age: 31-50; Gender: F)

- DV: Derivation of verb error: the verb form is not valid: it resembles the correct form but it has been incorrectly derived

Do you want to be **<#DV>prisoned | imprisoned </#DV>**? (Doc ID: 564614; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 1998; Age: 18; Gender: F)

Anyway, if we are together we can **<#DV> comeover | overcome </#DV>** any problem. (Doc ID: 581555; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: B2 Vantage; Year: 1999; Age: 20-30; Gender: F)

or the wrong part of speech has been used and is replaced by a verb

What do you think, can you **<#DV>advice | advise </#DV>** me? (Doc ID: 2646824; Nationality: Turkey; Exam: PET; CEFR Student Level Performance: A2 Waystage; Year: 2009; Age: 15; Gender: M)

- FV: A valid form of the verb has been used but it is incorrect in the context: Either the wrong choice has been made from the following: bare infinitive, to+infinitive, v-ing

I want **<#FV> sell | to sell </#FV>** a radio-CD Player. (Doc ID: 1094569; Nationality: Turkey; Exam: KET; CEFR Student Level Performance: A2 Waystage; Year: 2002; Age: 20-30; Gender: M)

I am looking forward to **<#FV> see | seeing </#FV>** your new report. (Doc ID: 562340; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: B1 Threshold; Year: 1997; Age: 20-30; Gender: F)

or an inflected form has been used when the bare infinitive, to+infinitive, or v-ing is required

They didn't **<#FV> mentioned | mention </#FV>** anything about the prices. (Doc ID: 567005; Nationality: Turkey; Exam: CAE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 1999; Age: 20-30; Gender: M)

How much does it **<#FV> costs | cost </#FV>** for two nights? (Doc ID: 569752; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: B2 Vantage; Year: 1997; Age: 17; Gender: F)

- IV: Inflection of verb error: A verb has been incorrectly inflected (often when irregular verbs are given regular infections)

My friend Dianna **<#IV> gived | gave </#IV>** it to me. (Doc ID: 1535179; Nationality: Turkey; Exam: KET; CEFR Student Level Performance: A1 Breakthrough; Year: 2004; Age: 12; Gender: F)

It **<#IV> costed | cost </#IV>** 450 dollars. (Doc ID: 2123756; Nationality: Turkey; Exam: KET; CEFR Student Level Performance: A2 Waystage; Year: 2007; Age: 12; Gender: M)

Her mother **<#IV> waked | woke </#IV>** her up. (Doc ID: 569954; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: B2 Wantage; Year: 1999; Age: 16; Gender: M)

Also **<#IV>helding | holding </#IV>** conferences related to diet … (Doc ID: 2251339; Nationality: Turkey; Exam: IAELTS AC; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 2008; Age: 20-30; Gender: F)

- MV   missing verb: a verb needs to be added to complete the phrase or sentence

I **<#MV> | am </#MV>** in Turkey now. (Doc ID: 575253; Nationality: Turkey; Exam: KET; CEFR Student Level Performance: A2 Waystage; Year: 1999; Age: 16; Gender: M)

Our last activities **<#MV> | are </#MV>** in July. (Doc ID: 562396; Nationality: Turkey; Exam: CAE; CEFR Student Level Performance: B2 Wantage; Year: 1999; Age: 20-30; Gender: F)

You **<#MV> | had </#MV>** better start thinking of purchasing the tickets from now. (Doc ID: 564265; Nationality: Turkey; Exam: CAE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 1998; Age: 20-30; Gender: M)

- RV   replace verb: a valid verb has been used but it is not correct in the context

She is **<#RV> named | called </#RV>** Sarah. (Doc ID: 762075; Nationality: Turkey; Exam: PET; CEFR Student Level Performance: A2 Waystage; Year: 2001; Age: 15; Gender: F)

Dear Sir/Madam I'm writing in order to **<#RV> have | get </#RV>** more information about your hotel. (Doc ID: 561798; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: B2 Wantage; Year: 1997; Age: 20-30; Gender: F)

The second mistake which is **<#RV> mentioned | made </#RV>** in your report is the name of the first group. The band were not called "Torrents" but "King Cup". (Doc ID: 563144; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 1997; Age: 20-30; Gender: F)

- TV: Tense of verb error: A valid inflected form has been used but it is not correct in the context (This code is not used when the corrected form is the bare infinitive, to+inf or v-ing. The code in that case is FV.)

But I **<#TV> don't change | haven't changed </#TV>** my bed. (Doc ID: 760382; Nationality: Turkey; Exam: KET; CEFR Student Level Performance: A2 Waystage; Year: 2000; Age: 13; Gender: F)

I **<#TV> also enclosed | have also enclosed </#TV>** the financial figures <#RV> of | for </#RV> the charity day in case you might need them. (Doc ID: 561802; Nationality: Turkey; Exam: CAE; CEFR Student Level Performance: B2 Wantage; Year: 2000; Age: 20-30; Gender: F)

I loved feeding the chickens, taking their eggs, <#RV> bringing | taking </#RV> them to my <#UP> grand-mother | grandmother </#UP> and observing with great attention the way she **<#TV> was cooking | cooked </#TV>** them. (Doc ID: 562757; Nationality: Turkey; Exam: CAE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 1998; Age: 20-30; Gender: M)

- UV: Unnecessary verb: A valid verb form has been used but its inclusion is incorrect

> I **<#UV> was | </#UV>** lost my dog. (Doc ID: 760451; Nationality: Turkey; Exam: PET; CEFR Student Level Performance: A2 Waystage; Year: 2000; Age: 20-30; Gender: M)

> I **<#UV> am | </#UV>** completely disagree with your report and would like to correct the errors. (Doc ID: 562340; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: B2 Threshold; Year: 1997; Age: 20-30; Gender: F)

> And the time was **<#UV>showing | </#UV>** exactly 2 o'clock! (Doc ID: 976821; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 2001; Age: 20-30; Gender: F)

3. Data Analysis

As a result of the analysis of the verb errors made by Turkish learners regardless of the proficiency levels and the error categories, we have found out that there are 16851 errors in the subcorpus of CLC consisting of 801.940 words. Table 2 presents the breakdown of these errors according to the verb error categories.
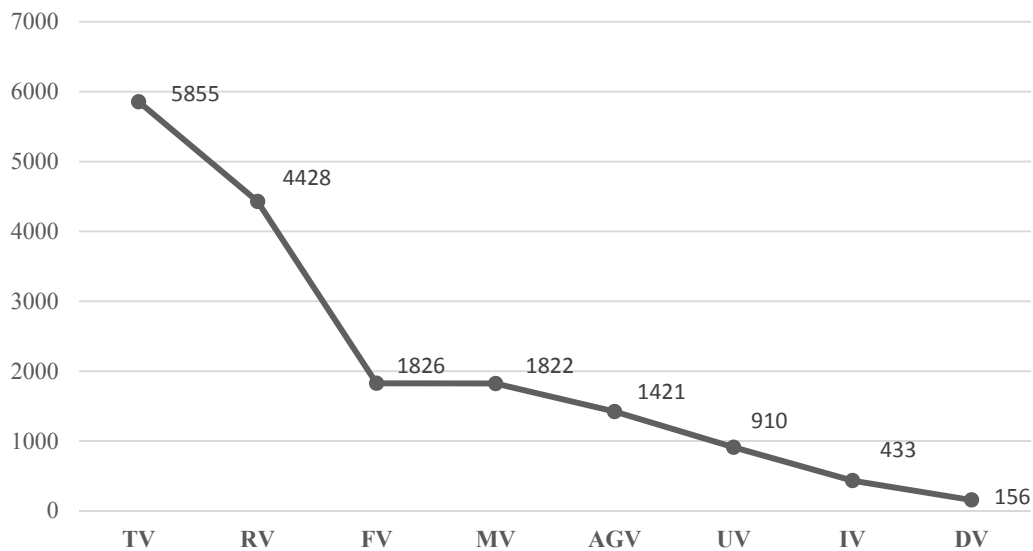
As displayed in Graph 1, the most common verb error category across the proficiency levels is related with the *tense of the verb* (TV). This error typology is followed by *incorrect verb choice in the context* (RV). These two most common error categories are followed by *incorrect verb form choice* (FV), *missing verb* (MV), *verb agreement* (AGV), *unnecessary verb* (UV), *verb inflection* (IV), and *derivation of verb* (DV) errors.

This graph presents a bird-eye view to the verb errors of Turkish EFL learners from A1 to C2 English language proficiency levels. In order to reveal the distribution of these error typologies according to each proficiency level of the learners, a further analysis has been carried out based on the CEFR levels.

Table 2. Distribution of Verb Error Typologies

CEFR Level	Corpus Size	Verb Error Typologies							
		TV	RV	FV	MV	AGV	UV	IV	DV
A1 – A2	289.888	2920	1517	658	860	441	468	230	24
B1 – B2	360.750	2331	2200	892	811	792	375	171	100
C1 – C2	151.302	604	711	276	151	188	67	32	32
Total	801.940	5855	4428	1826	1822	1421	910	433	156

When the breakdown of the verb errors according to the proficiency levels of the learners in Table 2 examined, it is inferred that the ranking of the verb errors do not change within the proficiency levels except RV errors in C1 - C2 level. As the table displays the number of the errors for each verb error category for each proficiency level and the corpus sizes are varied in each corpora, it is not possible to claim any statistical difference among the levels. For this reason, a log-likelihood (LL) analysis has been run to compare the error frequencies by normalising the corpus sizes. Table 3 displays the LL results for the verb errors between A1-A2 and B1-B2 groups.

Table 3. Log-likelihood results (A1-A2 vs. B1-B2)

Error Typology	Freq. in A1-A2	Freq. in B1-B2	LL	Significance
TV	2920	2331	257,67	0,000 +
RV	1517	2200	21,21	0,000 -
FV	658	892	2,78	0,095 -
MV	860	811	32,09	0,000 +
AGV	441	792	39,27	0,000 -
UV	468	375	40,68	0,000 +
IV	230	171	26,40	0,000 +
DV	24	100	34,91	0,000 -

According to the LL results, learners within the A1-A1 proficiency range make verb tense, missing verb, unnecessary verb, and verb inflection errors significantly more than B1-B2 proficiency level learners. The number of these errors decrease as the learners' proficiency level improves but incorrect verb form, verb agreement, and derivation of verb errors increases. When we compare B1-B2 and C1-C2 proficiency range learners (Table 4), we see that there is a significant decrease in all verb error categories in higher proficiency learners.

Table 4. Log-likelihood results (B1-B2 vs. C1-C2)

Error Typology	Freq. in B1-B2	Freq. in C1-C2	LL	Significance
TV	2331	604	121,67	0,000 +
RV	2200	711	38,09	0,000 +
FV	892	276	20,52	0,000 +
MV	811	151	100,10	0,000 +
AGV	792	188	54,99	0,000 +
UV	375	67	49,95	0,000 +
IV	171	32	20,90	0,000 +
DV	100	32	1,86	0,173 +

To reveal the persistent verb error categories across proficiency levels, learners in the A1-A2 and C1- C2 proficiency bands have also been compared. Table 5 displays the log-likelihood results in this regard.

Table 5. Log-likelihood results (A1-A2 vs. C1-C2)

Error Typology	Freq. in A1-A2	Freq. in C1-C2	LL	Significance
TV	2920	604	516,85	0,000 +
RV	1517	711	5,68	0,017 +
FV	658	276	9,56	0,002 +
MV	860	151	193,11	0,000 +
AGV	441	188	5,54	0,019 +
UV	468	67	132,88	0,000 +
IV	230	32	67,19	0,000 +
DV	24	32	12,17	0,000 -

As can be inferred from the table, the least significant differences among the verb error categories belong to *incorrect verb choice in the context* (RV), *incorrect verb form choice* (FV), and *verb agreement* (AGV) errors. It should also be noted that *derivation of verb* (DV) errors are significantly higher in C1-C2 level. Similar result is also observed in Table 4 presenting a low statistical difference in the same category of verb error between B1-B2 and C1-C2 levels. FV and DV errors might stem from the fact that the number of the verbs learnt increases in higher levels of proficiency but the rules pertaining to the correct use of bare infinitive, to+infinitive, v-ing forms and the derivation of the related verbs would still take some time to master as illustrated in the following examples.

> After that visit they decided to get on a ship **<#FV> enjoying | to enjoy </#FV>** the atmosphere of the Bosporus. (Doc ID: 563144; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 1997; Age: 20-30; Gender: F)

> This is the basic explanation for my dissatisfaction with the tour and I would like you to consider **<#FV> to refund | refunding </#FV>** some of my money to compensate me. (Doc ID: 563614; Nationality: Turkey; Exam: CAE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 1998; Age: 20-30; Gender: F)

> By doing so, you will **<#DV> practice | practise <#DV>** both, talking and understanding. (Doc ID: 764544; Nationality: Turkey; Exam: CAE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 2001; Age: 17; Gender: F)

As for the errors related with incorrect verb choice in the context, this can be considered one of the most persistent verb error category across the proficiency levels. Verb selection may be demanding for EFL learners because verb usage is highly related with the usage context and not consistently captured by local features. Learning how to use verbs in the context appropriately is challenging task as they often exhibit a variety of usages and each usage depends on a particular context. This error category is the highest second one in Turkish EFL learners' subcorpus in CLC. These delexical collocations in the corpus might be explained by L1 influence as some of them are highly frequent in Turkish. The learners might also be making these errors because they are still developing semantic and pragmatic competence in L2 to gain the knowledge that the collocability of lexical items requires more than semantic similarity. The most problematic verbs across levels are *make* (f: 121) and *do* (f: 112). For example:

> By using the Internet we can even **<#RV> make | have </#RV>** telephone conversations with a friend who is at the other end of the world. (Doc ID: 564383; Nationality: Turkey; Exam: CAE; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 1999; Age: 20-30; Gender: M)

> I would like to **<#RV> do | go </#RV>** shopping. (Doc ID: 1253658; Nationality: Turkey; Exam: KET; CEFR Student Level Performance: A1 Breakthrough; Year: 2003; Age: 12; Gender: M)

AGV errors are also one of the most persistent verb error typologies. Many studies conducted on AGV errors (Brown, 2000; Law, 2005; Stapha and Izaha, 2010) state that the most crucial factor is L1 transfer. On the other hand, some other studies (Moses, Joseph and Moses, 2007) claimed that these errors could also be connected with the fact that students acquire L2 morphemes in a fairly predictable fashion, regardless of language background and age, implying an interlanguage property. The most common AGV errors, which is related with subject-verb agreement, in Turkish EFL learners' corpus are related with the lexical verbs (f: 648) in present tense as presented in the following example:

> Especially in summer, the number of cars **<#AGV> increase | increases </#AGV>** because of the tourists. (Doc ID: 578491; Nationality: Turkey; Exam: CAE; CEFR Student Level Performance: B2 Vantage; Year: 1998; Age: 20-30; Gender: M)

This is followed by the agreement errors in auxiliary verbs: *is/are* (f: 400), *have/has* (f: 177), and *was/were* (f:171). For example:

> Do you think there **<#AGV> is | are </#AGV>** any tourist facilities around the hotel? (Doc ID: 562566; Nationality: Turkey; Exam: FCE; CEFR Student Level Performance: B1 Threshold; Year: 1997; Age: 20-30; Gender: F)

> In this competitive world the number of crimes **<#AGV> have | has </#AGV>** been increasing every day. (Doc ID: 2810852; Nationality: Turkey; Exam: IELTS AC; CEFR Student Level Performance: C1 Effective Operational Proficiency; Year: 2010; Age: 20-30; Gender: F)

> All sorts of creativity **<#AGV> was | were </#AGV>** praised. (Doc ID: 586070; Nationality: Turkey; Exam: CPE; CEFR Student Level Performance: C2 Mastery; Year: 1997; Age: 20-30; Gender: F)

4. Conclusion

This study has been conducted to reveal the verb errors made by Turkish EFL learners from A1 to C2 proficiency levels. According to the findings of the study, the most common verb error category across these proficiency levels is the incorrect choice of the tense of the verb. Incorrect verb choice in the context follows this. The other verb errors commonly made are incorrect verb form choice, missing verb, verb agreement, unnecessary verb, verb inflection, and derivation of verb according to their frequencies of occurrence in the corpus.

As can be inferred from the findings, the verb error rates decrease as proficiency level increases. Despite the decreasing tendency in the frequency of errors, some verb error categories are still persistent in C1-C2 proficiency band. Some of these persistent errors pertaining to *incorrect verb choice in the context*, *incorrect verb form choice*, *verb agreement* and

derivation of verb might stem from L1 transfer. The verb errors found in this study may also imply an interlanguage property but this requires a further in-depth contrastive interlanguage analysis involving the data from learners various L1 mother tongue backgrounds.

With regard to these errors made by the students along with A1-B2 proficiency range, in EFL classes, more emphases should be laid on inputting them lexical, grammatical and syntactic knowledge through data-driven teaching implementations, while as to those at higher English levels (C1-C2), lexical, especially idiomatic collocation knowledge is supposed to be emphasized. In order to do this, EFL teachers equipped with the skills of applying data-driven, corpus-based, teaching techniques, are advised to provide authentic materials as many as possible for students. When learners have more opportunities to be exposed to the conventionalized structures and idiomatic expressions, they are more likely to internalize them and utilize them in their language production. The more the EFL students have opportunity to receive the natural use of language input, the more they will be able to recognize and eventually produce structurally and semantically appropriate patterns.

Acknowledgement

References

Allerton, D. J., Tschichold, C., & Wieser, J. (Eds.), (2004). Linguistics, Language Learning and Language Teaching. Schwabe, Basel, Switzerland.

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of Syntactic and Morphological Accuracy by Advanced Language Learners. *Studies in Second Language Acquisition*, *11*, 17-34. https://doi.org/10.1017/S0272263100007816

Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

Brown, D. H. (2000). Principles of language learning & teaching. (4th ed.). New York: Longman. (pp. 49-58).

Chamot, A. U. (1987). The learning strategies of ESL students. In A. Wenden & J. Rubin (Eds.), Learner strategies in language learning (pp. 71-83). Englewood Cliffs, NJ: Prentice-Hall.

Corder, S. P. (1967). The significance of learner's errors. [Washington, D.C.]: ERIC Clearinghouse.

Corder, S. P. (1973). Introducing applied linguistics. Harmondsworth [Eng.]; Baltimore: Penguin Education.

Council of Europe. (2001). Common European framework of reference for languages: Learning, teaching, assessment. Cambridge, U.K: Press Syndicate of the University of Cambridge.

Cowan, R., Choi, H. E., & Kim, D. H. (2003). Four questions for error diagnosis and correction in CALL. *CALICO Journal, 20*(3), 451-463.

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. Elsevier, *26*(2), 163-174. https://doi.org/10.1016/S0346-251X(98)00001-3

Ellis, R. (1994). The Study of Second Language Acquisition. Oxford University Press, Oxford.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4*(2), 139–155. https://doi.org/10.1016/1060-3743(95)90004-7

Gentner, D., & Boroditsky, L. (2001). Early acquisition of nouns and verbs: evidence from navajo. Routes to language: Studies in honor of Melissa Bowerman, 32-5.

Gilquin, G., & Granger, S. (2015). From design to collection of learner corpora. *The Cambridge Handbook of Learner Corpus Research, 3*(1), 9-34. https://doi.org/10.1017/CBO9781139649414.002

Granger, S. (2003). The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL Quarterly, 37*(3). https://doi.org/10.2307/3588404

Grant, L. & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing, 9*, 123–145. https://doi.org/10.1016/S1060-3743(00)00019-9

Heift, T., & Schulze, M. (2012). Errors and intelligence in computer-assisted language learning: Parsers and pedagogues. London: Routledge.

Juozulynas, V. (1994). Errors in the Compositions of Second-Year German Students: An Empirical Study for Parser-Based ICALI. *CALICO Journal, 12*(1), 5-17.

Kellerman, E. (1983). Transfer and Non-Transfer: Where We Are Now. *Studies in Second Language Acquisition, 2*(1),

112-134.

Kızıl, A. Ş., & Kilimci, A. (2014a). Teaching Collocations through Web-Based Concordancing: In Tafazoli & Romero (Eds.), *Multiculturalism and Technology-Enhanced Language Learning, 4*, 142-156. http://doi.org/10.4018/978-1-5225-1882-2.ch009

Kızıl, A. Ş., & Kilimci, A. (2014b). Recurrent phrases in Turkish EFL learners' spoken interlanguage: A corpus-driven structural and functional analysis. *Journal of Language and Linguistic Studies*, *10*(1), 195–210.

Mitton, R. (1996) *English spelling and the computer.* Harlow, Essex: Longman Group.

Moses, G. O., Joseph, O. O., & Moses, V. O. (2014). Issues and challenges in teaching and learning of English in tertiary institutions: a case study of Kogi state college of education (technical), Kabba. *International Journal of Recent Scientific Research, 5*(11), 2003-2007.

Ndiaye, M., & Faltin, A. V. (2003). A Spell Checker Tailored to Language Learners. *Computer Assisted Language Learning, 16*(2-3), 213-232. https://doi.org/10.1076/call.16.2.213.15881

Olsen, K. A. (1999). An international perspective authors. Journal of the Association for Information Science and Technology.

Pantel, P., & Chklovski, T. (2004). VerbOcean: Mining the web for fine-grained semantic verb relations. *EMNLP, 4*, 33-40.

Schachter, J. (1974). An error in error analysis. *Language Learning, 24*, 205-214. https://doi.org/10.1111/j.1467-1770.1974.tb00502.x

Schuler, K. K. (2005). VerbNet: a broad-coverage, comprehensive verb lexicon. Penn Libraries, University of Pennsylvania.

Sinclair, J. M. (2001) 'The deification of information' in Thompson and Scott (eds), Patterns of Text, 287–313. https://doi.org/10.1075/z.107.14sin

Stapa, S., & Izaha, M. (n.d). (2010). Analysis of errors in subject - verb agreement among Malaysian ESL learners. *The Southeast Asian Journal of English Language Studies, 16*(1).

Swan, M., & Smith, B. (2001). Learner English: A teacher's guide to interference and other problems. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511667121

Taylor, B. (1975). The use of overgeneralization and transfer learning strategies by elementary and intermediate students of ESL. *Language Learning, 25*(1), 73-107. https://doi.org/10.1111/j.1467-1770.1975.tb00110.x