UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

Instituto de Ciencias Sociales y Humanidades

A CORPUS-BASED STUDY ABOUT ERRORS MADE BY LELI STUDENTS USING PLURAL NOUNS

Tesis para obtener el grado de

Licenciada en Enseñanza de la Lengua Inglesa

PRESENTA:

Garcia Villegas Maria Aurora

DIRECTORAS:

Dra. Abigail Carretero Hernández

Dra. Ana Abigahil Flores Hernández

Pachuca, Hgo. Mayo de 2026.

Universidad Autónoma del Estado de Hidalgo
Instituto de Ciencias Sociales y Humanidades
School of Social Sciences and Humanities
Área Académica de Lingüística

No. REF.UAEH/ICSHu/LELI/109/2026
Asunto: Autorización de impresión de tesis

DRA. ABIGAIL CARRETERO HERNÁNDEZ
COORDINADORA DE LA LICENCIATURA EN ENSEÑANZA DE LA LENGUA INGLESA
P R E S E N T E.

Manifestamos a usted que de conformidad con el Título Cuarto, Capítulo I, Artículos 37, 40 y 41 del Reglamento de Titulación vigente de nuestra universidad, se autoriza la impresión formal del trabajo de investigación de la pasante **María Aurora García Villegas** con número de cuenta **375703** bajo la modalidad de Tesis Individual cuyo título es ***"A corpus-based study about errors made by LELI students using plural nouns"*** debido a que reúne los requisitos de decoro académico a que obligan los reglamentos en vigor para ser discutidos por los miembros del jurado.

Miembros del jurado	Cargo	Firma de aceptación del trabajo para su impresión formal
Dra. Hilda Hidalgo Avilés	Presidente	
Dra. Abigail Carretero Hernández	Secretario	
Dra. Ana Abigahil Flores Hernandez	Vocal	
Mtro. Tomás Hernández Ángeles	Suplente	

C. c. p. Archivo
HAH/ACH/ghe*

Carretera Pachuca-Actopan Km. 4 s/n, Colonia San Cayetano,
Pachuca de Soto, Hidalgo, México; C.P. 42084
Teléfono: 771 71 7 20 000 Ext. 41024, 41023
jaali_icshu@uaeh.edu.mx

Index

Chapter 1. Introduction

The nominal and verbal inflectional morphemes, which are crucial for grammatical accuracy and linguistic proficiency, often present unique challenges for learners due to their complexity and variability across contexts. The analysis of vocabulary patterns in a spoken corpus of Mexican L2 English provides valuable information about the lexical choices, frequency, and usage trends of Mexican learners and the challenges they face when speaking in a foreign language.

Despite the importance of inflection patterns in the learning of English as a Foreign Language (EFL), which according to Lieber (2009), is the ability to create new lexemes versions so they can fit into a sentence context, there are just few studies and teaching strategies focused on the learners' needs related to these two types of inflection and the errors made. For example, Demarta (2012), Maslen, Theakston, Lieven, and Tomasello (2004), Ying and Yang (2022), and Minkkinen (2015) carried out studies with Spanish, Cantonese, Japanese and Arabic L1 learners to analyse inflection patterns.

Thus, the aim of this study is to analyze patterns of nominal and verbal inflection in the vocabulary learning in learners' oral production of English as a foreign language because there is a significant lack of strategies designed to effectively lead the diverse learning needs of students when it comes to mastering inflected morphemes.

For the purposes of this study, the data was gathered from the Mexican Learner Corpus (Flores & Moore, 2023) and it was analyzed using the software UAM Corpus Tool (O'Donnell, 2008) in order to categorize the errors made by learners through layers based on the classification of Akbas and Dincer, 2021; Ardébol, 2012; Mahmood, Mahmood and Rasool, 2020 and Fontana, 2013. The tags used are target-like use, non-target-like use, which is divided into subcategories such as underuse, misuse (misselection, misrealisation), overuse, and unclassified. In this way, the relevance of this study relies on its contribution to the field

as a result of the lack of corpus-based research that explores the specific linguistic errors made by Mexican speakers studying for a degree in English Language Teaching.

This study is organized into six chapters, each of which addresses key components of the research. Chapter one provides an introduction to the concept of inflection, its paradigms, the emergence and accuracy criterion, and an overview of relevant studies that utilize corpus methodology to examine and analyze inflectional morphemes. This foundational chapter presents the theoretical framework for the study. Chapter two focuses on the learner corpus, including a detailed discussion of the MexLeC corpus, which serves as the primary data source for this research. Chapter three outlines the methodology employed in the study, describing the research design, data collection procedures, and analytical approaches. Chapters four and five present the results and discussion, respectively, delving into the findings and their implications in detail. Finally, Chapters six and seven are dedicated to the conclusion, summarizing the study's key contributions.

Chapter 2. Literature Review

2.1 Inflection

Morphology, according to Lieber (2009), is the study of word formation, including the ways new words are coined in the languages around the world, and the way words are varied depending on how they're used in sentences. In addition, Haspelmath (2002) states that morphology is the mental system involved in word formation and in other words, it is the branch of linguistics that deals with their internal structure and how they are formed. In this way, the definition taken for this study is that morphology is the branch of linguistics in charge of studying the word formations that a word may have.

There are mainly two types of morphology: derivational, which is the change of world class of the base lexeme. For example, nouns can be derived from verbs, adjectives from nouns, and the list can continue because the base morpheme changes its grammatical category (Haspelmath, 2002). Meanwhile, the inflection in the word process occurs when the base morpheme does not change its grammatical category; this concept will be explained deeper in the next section.

According to Lieber (2009), inflection is a process of word formation where the morpheme does not change of category and does not create new lexemes, but rather changes the form of lexemes so that they fit into different grammatical contexts. Moreover, this category of word formation can provide grammatical meaning which includes information related to the number such as singular or plural, person like first, second or third, tense that could be past, present or future. Haspelmath (2002) states that the inflectional categories do not have a clearly identifiable meaning, but only a syntactic function that differs hardly at all in function. Furthermore, when different markers express an inflectional category, this is called an allomorph.

Aronoff and Fudeman (2011) define inflection as the realization of morphosyntactic features through morphological means. This means that the lexemes express morphosyntactic information that includes syntactic categories such as past, plural, imperfective, and genitive; these syntactic categories are frequently related to morphosyntactic features or morphosyntactic properties. Moreover, McCarthy (2002) states that the inflected forms of words are the type of variation that words exhibit based on their grammatical context. Furthermore, Booij (2005) mentions that inflection is the morphological marking of properties on a lexeme, resulting in a set of grammatical lexemes with several forms for that lexeme. What is more, there are two main inflectional dimensions for nouns that are found in many languages and those dimensions are number and case; these dimensions are referred to morphosyntactic categories because they play a role in both morphology and syntax. Additionally, some examples of the morphosyntactic features of these two dimensions are in the case of nouns, singular or plural, and verbs are usually inflected for categories such as tense, aspect, and mood.

Finally, inflection is a word formation process where a morpheme neither changes the word's grammatical category nor creates a completely new word. It is important to mention that each lexeme can provide different information according to the type of morpheme which can be either nominal or verbal as in the case of nouns, they can provide information such as number, case, gender and noun class, while in verbal inflection can be found information as person, case, tense, aspect, mood and modality.

2.2 Inflectional Paradigms

According to Lieber (2009), Aronoff and Fudeman (2011), Booji (2005), McCarthy (2002), and Haspelmath (2002), there are three primary types of inflectional paradigms: nominal, verbal, and adjectival. The nominal paradigm is responsible for modifying nouns,

such as in the use of pluralization or possessive forms. The verbal paradigm involves changes to verbs, including tense, aspect, mood, and agreement with the subject. Lastly, the adjectival paradigm modifies adjectives, as seen in the formation of comparative and superlative degrees. However, this research narrows its focus to two specific types of inflection: nominal and verbal, which are the most common due to the tasks from the corpus; adverbial inflection was rarely observed in the transcription. These paradigms are considered most relevant to the study's objectives and are explored in greater depth in Sections 1.2.1 and 1.2.2

2.2.1 Nominal Inflection

Nominal inflection refers to the various word formation processes that occur with nouns. It includes several types of inflection, such as number, gender, noun class, and case. These features modify the form of a noun without changing its grammatical category, allowing it to convey additional grammatical information.

One of the most familiar inflectional categories for English speakers is number, where nouns are marked as singular or plural. For example, in singular nouns, the words cat, mouse, ox, and child, to mention some words, and the plural of the same words are cats, mice, oxen, and children. Generally, in English nouns are pluralized by adding -s; however, some nouns form their plurals irregularly (Lieber, 2009). According to Aronoff and Fudeman (2009), in number it can be found the singular and plural number of nouns, in less frequency and in other languages there exists a dual number which refers to two items, trial number refers to a set or three items, and finally the paucal number, which comes from the Latin *pauca*, meaning few. In addition, Booij (2005) expresses that the option of a special number is determined by the information that the speaker wants to express, and this is called inherent inflection.

Gender is another inflectional category that is frequently used in languages such as French, Spanish, German, Latin, Russian, and other Indo-European languages. In different

languages, the gender masculine and feminine genders, and sometimes a neuter. Furthermore, grammatical gender is sometimes determined by the biological sex of the noun's referent, depending on the semantic class to which the noun belongs. Nevertheless, in many instances, gender assignment occurs with a considerable degree of arbitrariness. However, many words are assigned grammatical gender in accordance with their natural gender, reflecting whether the referent is perceived as masculine or feminine.

In addition, Lieber (2009) describes cases as another category that is used to inflect nouns. This category is useful to determine how the noun is distinguished on the basis of how they are employed in sentences; for example, this inflectional category helps to identify whether it functions as a subject, direct object, indirect object, as a location, time, or instrument. In addition, Aronoff and Fudeman (2011) state that there are more syntactic functions, such as the ergative with and absolutive case engaged to different objects that can be either transitive objects or intransitive subjects; on some occasions, it may include some notions, such as the locative focusing on the place or instrumental. What is more, according to Booij (2005), all the case marking depends on the syntactic context in which they appear, and this phenomenon is known as contextual inflection.

2.2.2 Verbal Inflection

Verbal inflections convey important grammatical information such as person, tense, aspect, mood, case, and voice. This implicit information is essential for transmitting a complete message. As with nominal inflection, these types of verbal inflection do not alter the grammatical category of verbs.

Person is a category that is used to inflect verbs; it is basically when the verbs have different endings depending on whether the subject is in the sentence. For instance, when the subject is the speaker is called first person; when she or he is the hearer, the subject is acting as second person, and finally, the subject is a third person when he/she is acting as someone

else. Additionally, another kind of person marking is the distinction between the inclusive and exclusive forms of the first-person plural; in an inclusive form, the speaker includes herself and the hearer, while in the exclusive form, the speaker includes herself and others, but not the hearer.

As previously mentioned in the section on nominal inflection, both nouns and verbs share the inflectional category of case. In the case of verbs, this is reflected in their transitivity: verbs are considered transitive when they take an object, and intransitive when they do not (Lieber, 2009).

Tense and aspect are inflectional categories that express temporal features of verbs, each in a distinct manner. Tense situates an event in time relative to a reference point, which is typically the moment of utterance. The three primary tenses in English are present, past, and future. The present tense indicates that the event coincides with the time of speaking, the past tense denotes that the event occurred before the moment of speaking, and the future tense signifies that the event will occur after it.

In English, the suffix -ed is used to talk about past events using regular verbs; nevertheless, there are no suffixes for the future, instead an auxiliary verb needs to be used, in this case, *will,* which is called periphrastic marking. Then, aspect conveys information related to the internal composition of the event or the way in which the event occurs in time. Two main aspects are included within this inflectional category, the first is the **perfective aspect**, which presents an event as completed, that is, it is viewed from an external perspective, and its internal temporal structure is not relevant. The second one is the imperfective aspect, and here the event is viewed as ongoing, and it is observed from inside. In this way, other types of aspects focus on the particular points of an event as the inceptive, which pays special attention to the beginning of an event, the continuative aspect is the middle of the event in progress, and the completive aspect is the end of the event.

These inflectional categories are related to the speech acts expressed by a verb. There are three moods of speech acts: declarative, used to make statements; interrogative, used to ask questions; and imperative, used to give commands. Another distinction in mood and modality is between a realistic form, which the speaker uses to signal that the event is actual or can be verified through perception, and an irrealistic form which indicates that the event is imagined or conceived as possible. Nevertheless, according to Aronoff and Fudeman (2011), in English, modal verbs can incorporate **may** and **must**, which both express different degrees of compliment, the first one tells obligation, meanwhile the second one has to do with truth.

Finally, voice is an inflectional category that allows for focusing on different noun phrases in sentences. There are two types of voices: the active voice, in a sentence with an agent and patient, and the passive voice, in which the patient is the subject of the sentence. In this case, the passive in English is expressed periphrastically by a combination of the auxiliary verb to be plus the past participle (Lieber, 2009). Additionally, in the words of Aronoff and Fudeman (2011), the most usual difference between these two forms of voice is that the passive he subject is the patient, contrary to the active.

2.3 Relevant Studies on Morphological Inflection

There have been different studies done around the world, such as Yin and Yang (2022), who made a corpus-based longitudinal study to provide new evidence concerning the acquisition of English plural making by a Cantonese English simultaneous bilingual child and an English-speaking monolingual child. As part of the methodology, the informants were drawn from two different corpora: the CHILDES Cantonese-English Yip/Matthews and the CHILDES English MacWhinney Corpus.

Then, the utterances containing noun phrases in obligatory plural context were extracted from the corpora; after that, they were organized on an Excel sheet. The coding

process consisted of first determining whether the noun was regular or irregular plural; second, noting the addition of a regular plural ending to nouns; and third, checking instances of double marking in irregular nouns. The error rate was calculated by dividing the number of errors by the total number of obligatory plural contexts. The results showed that the bilingual child produced more errors in plural making, with all of them being required but omitted (RO) errors; in the same way, it was found that over-regularisation was absent in the bilingual's production of plural marking at a later age. In conclusion, it was found that the bilingual child lagged behind the monolingual child and achieved mastery of plural marking at a later age.

Ardebol´s (2012) study aimed to incorporate the analysis of data taken from a corpus compiled from secondary school learners of English, within the framework of morpheme order studies. He used a learner profile, an elicitation task, and a placement test to compile the list of texts to produce the corpus, then tagged the morphemes using UAM Corpus Tool. Subjects in their study were 142 students holding the CEFR levels from A1 to C1 from two junior high schools: Instituto de Educación Secundaria (I.E.S) San Juan de la Cruz and from the I.E.S Santisima Trinidad.

Participants performed an elicitation task called “frog, where are you?” and these were transcribed and analyzed through UAM Corpus Tool software with the tag set of target-like use, non-target-like use subdivided into underuse, overuse, and misuse, divided by misselection and misrealization. The results showed that, among A2 students, the most accurate morphological category was the lexical be, with an accuracy rate of 0.69, followed by plurals (0.60), articles (0.56), past irregular forms (0.46), auxiliary be (0.41), the progressive -ing (0.39), past regular forms (0.32), and finally the third person singular form (0.09).

Similarly, Demarta (2012), using Ardébol´s methodology, aimed to find out if there were similarities between their participants' morpheme acquisition and the pedagogical implications that a specific morpheme acquisition order can imply. This study was conducted in two Secondary High Schools with students between 12 and 19 years old from Granada. The writer employed four principal tools: the learner profile form, a placement test, an elicitation task, and the UAM Corpus Tool. In general, the results showed that the morphemes with more target-like use were articles, lexical be, plurals, progressive, possessives, the verb to be as auxiliary, past regular, past irregular, and third person of present singular. Finally, it was concluded that students who have gotten a high proficiency level use the plural forms more frequently and correctly than those who are at beginning levels.

Furthermore, Fontana (2013), as in the previous two studies, Ardebol's and Demarta's, Fontana followed the same methodology and the same tag set to use in the UAM Corpus Tool. She aimed to shed light on the acquisition of inflectional morphology in EFL learners whose first language is Spanish. The learner corpus used in this research was collected from a total of 79 English levels, from the starter to the advanced proficiency level. Fontana found that lexical be, progressive, and plurals are the morphemes with the highest percentages with target-like use, followed by the third singular -s morpheme, and with less target-like use is past regular and irregular.

In addition, Maslen, Lieven & tomasello (2004) conducted a corpus study on past tense and plural overregularization in English. For this study, a computerized language analysis (CLAN) was used to extract all child utterances containing correct and regularized past tense forms of irregular main verbs, as well as correct and regularized plural forms of irregular nouns. The results showed that of 19 irregular plural nouns, 10 were overregularized at least once. In addition, 1,263 past-tense tokens were produced, where 52 irregular verbs and 24 irregular verbs were overregularized.

Moreover, Akbaş and Dinçer (2021) highlighted that the fixed natural order of grammatical morphemes was established through manual analysis. This analysis was conducted using data drawn from an EFL corpus; in this study, the author chose the exam scripts of students in an English language teaching department. Students answered two different timed argumentative writing tasks, and the corpus selected consisted of 136 exam scripts, but the authors selected just 68 students' exam scripts. The scripts were manually analyzed through the UAM Corpus tool using the tags of conform target-like and non-target-like use, subdivided into underuse, misuse, and overuse. where the inflectional morphemes with the greatest target-like use are plural-s, lexical and auxiliary be, third person singular -s, and at the end, the past irregular and regular. Results reported that some morphemes that were ranked as low in comparison with the so-called natural order were almost non-existent features in participants' mother tongue.

Altarawneh and Hajjo (2018) analyzed the extent to which Arabic-speaking EFL learners are aware of the English plural morphemes and whether they can recognize them in context. The participants in this study, conducted at A1 Ain University of Science and Technology in the United Arab Emirates, were divided into two groups: 30 beginner learners and 30 intermediate learners. As an instrument, they used a Grammaticality Judgment Task (GJT) in which sentences used in the test were adapted and modified from the Corpus of Contemporary American English (COCA) in order to suit the students' English proficiency level. The results showed that there is little awareness of the English plural morphemes among Arabic-speaking EFL learners.

Minkkinen (2015) studied the influence of morphological congruency on the acquisition of the English plural morpheme -s. The study aimed to lend support for language transfer theories and find any indications for conceptual transfer. The corpora used were the ICLE (International Corpus of Learner English), which consists of written data, and the

LINDSEI (The Louvain International Database of Spoken English Interlanguage which is made of spoken interviews. The author looked for correct and incorrect plurals and the total words in both corpora. The results revealed that even the qualitative analysis failed to provide any definitive evidence on the effect of conceptual transference on the L2 acquisition, leading to support for the still very small amount of research conducted on the subject. The study concluded that Japanese learners were significantly more likely to produce plural errors in both their writing and speech.

Moreover, Khor (2012) investigated the morpheme acquisition order of Swedish students in grades 6 and 7, utilizing corpus texts drawn from the Uppsala Learner English Corpus (ULEC). This study examined three morphemes: articles, the preposition in, and the plural form. The global results showed that the errors that both made were consistent with the errors that were found in Khor (2012). As part of the methodology, the samples were processed manually two times, just using pen and paper, in order to check that there were no mistakes. It is important to mention that the morphemes from each text were documented in a matrix to measure the frequency. The tags that he used in order to analyze the data were omission, addition, misinformation, and misordering. A phenomenon called overgeneralization was found in terms of the -s that is attached to irregular nouns, while the plural -s is not a problem for Swedish learners.

On the other hand, Zhang and Widyastuti (2010) made and studied the acquisition of second language morphology based on the emergence criterion. The aim of this study was to assess the acquisition outcome of six English morphemes as *past tense -ed, plural -s, (stand-alone), verb-ing form, aux+ verb forms including progressive be+ verb-ing, perfective* and *have+ verb-en*; in the interlanguage of three members of an Indonesian family one year after they arrived in Australia. The participants were three members of a family from Indonesia composed of parents and their 5-years-old daughter where the father is the most

advanced and the daughter the least. In this way, the main goal of this research was to find out how much English morphology the informants have acquired after one year of living in the English-speaking environment. For this research the speech data was collected individually through oral interviews at the informants's home; this interview followed a structured format, with the interviewer providing talking points and tasks specifically to elicit the targeted grammar forms mentioned previously. The results showed that all the target English morphology emerged in the L2 of the father, the lexical and phrase plural -s emerged in the L2 of the mother, the lexical plural -s emerged in the L2 of the daughter, the verb-ing form was used only by the daughter and the verbal phrases instances were not sufficient in the data of the mother and the daughter.

In conclusion, the studies reviewed provide important insights into the acquisition of English grammatical morphemes among both first and second language learners. Overall, the findings show that the development of English morphology is gradual and influenced by several factors, such as learners' proficiency level, first language background, and the amount of exposure to the language. Across different contexts and methodologies, researchers have observed that some morphemes tend to be acquired earlier and with greater accuracy than others, while errors such as omission, overregularization, and misuse are common during the learning process.

2.4 Criteria in the Learning of Morphology

In this study of morphological acquisition, two key criteria are used to evaluate learners' development: emergence and accuracy. The emergence criterion focuses on the initial appearance of a morpheme in the learner's language. According to Pallotti (2007), a morpheme can be considered to have emerged only when it appears in more than four instances, indicating consistent use rather than isolated occurrences. In contrast, the accuracy

criterion assesses how correctly a morpheme is used once it has emerged. Fontana (2013), Ardébol (2012), and Demarta (2012) argue that a morpheme is considered learned when it is used accurately in at least 90% of obligatory contexts. These criteria provide a comprehensive framework for determining both the initial emergence and the stable acquisition of morphological forms.

2.4.1 Emergence Criterion

As it is known, Krashen and Terrell (1983) said that there are five stages of second language acquisition, where the first one is the preproduction in which a student has minimal comprehension, does not verbalize, and nods yes and no. Secondly, the early production is present where a student has limited comprehension, he/she produces one- or two-word responses and participates using key words and familiar phrases, and the students use only present tense verbs. After that, speech emergence is when the student has good comprehension, can produce simple sentences, and make grammar and pronunciation errors. Then, in the intermediate influence stage, the student has excellent comprehension and makes few grammatical errors. Finally, in the advanced fluency students have a near native level of speech.

The organization and interpretation of data in morphological learning research involve a series of systematic stages following the collection and transcription of information. To ensure accurate analysis and interpretation, the data must be organized appropriately. A crucial initial consideration is whether the analysis will be based on types, tokens, or both. In this context, *types* refer to distinct morphological categories under examination, while *tokens* represent all instances, including repetitions, of those types as identified according to specific criteria. Secondly, researchers must determine whether to include all items in distributional tables or to exclude certain data points for methodological reasons (Pallotti, 2007).

Data interpretation, particularly through distributional tables, involves identifying patterns that represent the first systematic and productive use of a given morpheme or structure. Determining whether a morpheme has emerged requires examining three key aspects. The first one is data robustness, which refers to the minimum amount of evidence needed to support claims of emergence. According to Pallotti (2007), a morpheme must appear in at least four distinct instances to satisfy this criterion and to operationalize the construct of emergence. The second aspect is productive use, where evidence must show that the morpheme is used in a way that extends beyond rote repetition. This includes three subcategories: (1) the presence of morphological minimal pairs (e.g., singular and plural forms in appropriate contexts), (2) overgeneralizations, where learners apply a rule too broadly or create novel forms, and (3) sufficient lexical variety, indicated by the use of the morpheme with multiple different stems or affixes.

The third and final aspect is systematic distribution, which requires that a sufficient number of tokens appear across a wide variety of contexts, demonstrating both productivity and generalization. This aspect is closely tied to the notion of *suppliance in obligatory contexts* (SOC), as a morpheme can only be considered to have emerged if it consistently appears in appropriate grammatical environments. Here, the concept of *target-like use* becomes relevant, indicating that the morpheme is not only present but correctly used in contexts where it is required.

2.4.2 Accuracy Criterion

Housen and Kuiken (2009) define accuracy simply as “error-free” speech. In terms of inflection, often measured by the learner’s suppliance with a specific form in obligatory contexts (i.e., where a plural, progressive or past tense form is needed). The researcher usually decides which form and context to measure based on a developmental inquiry

(proficiency) or on task conditions. In many cases, studies involving lower-proficiency learners assess accuracy based on different target forms than those used in studies with higher-proficiency learners because accuracy on initial target forms is expected to approach ceiling levels as proficiency increases.

However, if accuracy is determined by correctness of certain selected forms only, accuracy and development (perhaps complexity or proficiency) are confounded in the measure and potentially misleading. For this reason, developmental studies on the acquisition of grammar patterns have tended to measure accuracy. Four approaches to calculate such accuracy rates are holistic scales, error-free units, number of errors and target language use formula.

The holistic scale is used to assess linguistic or grammatical accuracy as one component among others in a composition rating scale. In this scale, the independent variable is the type of task which is in charge of measuring accuracy in grammar and vocabulary mechanics and the independent variable is the grade level and age of arrival, influencing accuracy. In the Error-Free T-units (EFT) methodology, the independent variable is the time and the accuracy measure is the percentage of EFT words per EFT considering as the independent variable the teaching task, which can be guided through answering questions or free picture description. Calculations for the accuracy measure in this case are the following: the ratio of EFT divided by total T-units ratio of EFT, or the total of clause words in EFTs divided by total words.

Regarding the number of errors, Carlisle (1989) measures accuracy by taking the average number of errors per T-unit (sentence), classifying them into mechanical, lexical, morphological, phonological, and syntactic errors. As an alternative procedure, Zhang (1987) measured accuracy by computing the number of errors per 100 words, where the only independent variable is the cognitive complexity of the question and response.

In terms of pure inflectional analysis accuracy, in 1973, Brown introduced the term Suppliance in Obligatory context (SOC) to establish not only the importance of the inflectional output, but the importance of the "where-is-required" output and stated the mandatory 90 per cent to establish accuracy criterion acquisition of an inflectional pattern. This rate of accuracy is calculated after Pica´s (1988) and Stauble (1981) guidelines as follows: Target Language Use (TLU), the measurement of accuracy, is obtained by dividing correct Suppliance in Obligatory Context (SOC) by the sum of Obligatory Contexts (OC) plus the Suppliance in Non-Obligatory Contexts (SNOC). Considering SOC, the cases of correct use of inflectional patterns when required; OC the same number of cases in which inflection was required (correctly provided or not); and, SNOC the cases in which an inflectional pattern was produced without being required. For the purpose of this research, the study of the inflectional patterns, TLU formula, are considered.

2. 2 Learner Corpus

2.2.1 Corpus linguistics

McEnery and Hardie (2011) defined corpus linguistics not as the study of a particular aspect of the language instead of that the author defined it as an area that focuses upon a set of procedures or methods to study a language and introduced two types of studies based on corpus linguistics. The first one is, according to Tognini-Bonelli (2001), the corpus-based approach, in which corpus data is typically used to explore a theory or hypothesis that is previously established in literature, and the researcher aims to validate, to refuse or to refine that theory. On the other hand, the second type is the corpus-driven approach, where the corpus itself serves as the primary source of language hypotheses.

McEnery, Xiao and Tono (2006) narrates that at the first signs of corpus linguistics research, researchers tended to use paper slips rather than computers as a means of data

storage, and the term corpus made emphasis on simple collections of transcribed texts and thus was not representative. As it was mentioned above, the corpus methodology was based only on paper-based corpora to study grammar through paper slips, human hands, and eyes. However, the development of technology, in specific computers, has facilitated this study with their increasing power and massive storage at relatively low cost.

2.2.2 Learner Corpora

According to Granger (2008), learner corpora emerged as a consequence of the technological advances in and how they helped to computerize corpus linguistics and how this can benefit applied linguistics. Learner corpora often prioritize learner populations categorized by their first language background. This focus is largely due to the product-oriented nature of the methodology, which emphasizes observable language performance rather than underlying competence. Moreover, this approach upholds a central principle of corpus linguistics: authenticity. By allowing learners to express themselves freely, rather than responding to prompts targeting specific linguistic features, the data collected better reflects natural language use.

In addition, Callies and Paquot (2015) defined learner corpora as systematic collections of authentic, continuous and contextualized language use, which can be either spoken or written, by second language learners stored in electronic format. This format is usually a special type of empirical data used by scholars in a variety of disciplines. This means that it is a compilation of data used to study how students are using the second language through an electronic format; the language analyzed needs to be spoken or written. Another definition of learner corpus given by Granger (2008) is that it is a linguistic methodology that was founded on the use of electronic collections of naturally occurring text corpora. This new linguistic methodology has been used for the last two decades.

Finally, Granger (2008) defined learner corpora as electronic collections of texts produced by language learners. Moreover, the implementation of learner corpora in the research field has contributed to the study of the second language acquisition theory, provides a better description of interlanguage and a better understanding of the factors that influence it, also contributes to covering the language needs as the development of pedagogical tools and methods.

Learner corpora can be classified in attention to its different characteristics such as size, modality, use and time. Regarding their size, corpora can be big or small. For example, the Trinity Lancaster Corpus (TLC, Trinity College London, 2025) contains approximately 4 million tokens, while smaller corpora such as the Mexican Learner Corpus (MexLeC, Flores & Moore, 2024), with around 400,000 tokens, illustrate what is considered a small corpus. However, size is not only about the number of tokens, but also about the generalisability or representativeness of the results. For this reason, a small corpus can still be of considerable value.

Modality refers to the type of discourse produced by learners: whether written, spoken, or both. Using the same examples as before, we can say that TLC and MexLeC are spoken corpora, as they contain audio or video recordings that have been transcribed into text. In contrast, the International Corpus of Learner English (ICLE, Granger et al., 2020) is an example of a written corpus, based on essays produced by intermediate to advanced learners. Finally, we also find multimodal corpora, such as the Santiago University Learner of English Corpus (SULEC, 2025), which includes two components: semi-structured oral interviews, and compositions and argumentative essays.

According to time, a corpus can be longitudinal or cross-sectional, longitudinal corpora refers when the data that comes from the same learners and is collected over time, and cross-sectional corpora contain data gathered from different categories of learners at a

single point in time. Quasi-longitudinal corpus is gathered at a single point in time, but from learners of different proficiency levels. An example of a longitudinal corpus is MexLeC, which contains data gathered from university students who complete the instrument (an interview) three to four times, once per year during their bachelor studies. In contrast ICLE, a non-longitudinal corpus, collects data only once. Finally as an example of quasi-longitudinal corpora we have TLC and SULEC also gathering data at a single point in time, but from learners at different proficiency levels. A2 to C2 in the case of SULEC, and B1 to C2 for TLC.

The last category concerns whether the database is intended for immediate or delayed pedagogical use. The second type, delayed pedagogical use, refers to corpora that are not used directly as teaching or learning materials by the learners who produced the data. Instead, the purpose is to provide a better description of a specific interlanguage and to design tailor-made pedagogical tools that may benefit similar learners. In contrast, for immediate pedagogical use, the data are collected by teachers as part of learners' daily classroom activities. It is important to note that this type of corpus is usually not as representative as other corpora. According to this distinction, we could consider MexLeC and SULEC as corpora collected to enhance knowledge about local learners and therefore for immediate applications. In contrast, ICLE would fall under delayed applications, since it has been mainly used for research purposes.

For this research, the concept of learner corpus is the following: a systematic collection of computerized data in which the language needs to be authentic, contextualized and continuous in order to analyze how learners used the second language. The approach taken will be corpus-based because, according to the literature, it looks to confirm or refute different hypotheses to provide evidence about how the learner uses the language in a natural context.

2.2.3 Mexican Learner Corpus

The Mexican Learner Corpus (MexLeC) is a learner corpus created by Flores and Moore (2023), is a longitudinal collection of oral texts produced by learners of English from Mexican universities. This corpus aims to provide a basis for research on foreign language learning in Mexican learners and for future applications in the design of strategies and materials for English language teaching.

All the participants are university students of a bachelor in modern languages (translation and/or teaching), their data about gender, age, mother tongue, and foreign language learning background and L2 proficiency level are contained in the data section.

To elicit data a 16-minute oral interview divided into four tasks was designed. The interaction in this interview is one-on-one (interviewer and participant). All the interaction is video-recorded via Zoom, which allows for recording the interviews, then they are uploaded to the well-known platform Google Drive to preserve them.

Tasks in the interview are mainly four, where the first task is open-ended questions on familiar topics such as family, friends, occupation, and leisure activities, and if they already have plans with friends, family, or in their job or school; in this part, there are four questions of one minute each about just two topics.

The second task is about the memories that they have about the previous two topics selected in task number one which can be about free time, family, friends and work or school in each narrated memory the participants have one minute of each one in total of two minutes in this first section of the task; on the other hand, the second part of the task is a picture-based narrative in which the participant has to narrate a story based on eight images that are presented in a total of two minutes.

The third task is questions about likes and preferences, with some options given in which the participant has to make choices and support them. The last task is opinion questions on social, educational, and/or technology issues, where two types of questions are selected and the participants have just two minutes to answer each question in a total of four minutes. In this way, each interview from the corpus must last sixteen minutes.

The transcription guidelines for interviews have been adapted from The Trinity Lancaster Corpus (Trinity College London, 2025) and the LINDSEI Corpus (Center for English Corpus Linguistics, 2021). Information on the profile of the participants, such as learning experiences, mother tongue, and their contact with other foreign languages, has been collected to help the data interpretation of the corpus. The interviewers are Mexican native Spanish speakers holding the levels of proficiency B2 and C1.

All in all, the Mexican Learner Corpus represents a valuable contribution to the study of English as a foreign language in Mexico. Its longitudinal design, systematic data collection, and the participants´ profiles provide researchers to analyze learner development over time. Furthermore, by combining authentic oral production with carefully structured interview tasks, the corpus offers meaningful insights into the language use, communicative competence, and learning processes of Mexican university students. This resource not only supports empirical research but also has practical implications for the improvement of language teaching practices and material design in the Mexican educational context.

Chapter 3. Methodology

In this chapter the methodology for this research is explained in detail, providing a comprehensive overview of the type of research being conducted. This explanation includes an analysis of the nature of the data being utilized, the timeframe for data collection, the depth and scope of the research, and the specific sources of the data. Furthermore, this chapter addresses the core research questions that guide the study. These questions aim to investigate the inflectional patterns that emerge in the oral production of Mexican English learners. Specifically, the research seeks to identify which inflectional patterns occur most frequently and to explore the developmental sequence through which these patterns are learned.

According to the nature of the data, this research is a mixed-method approach, Creswell and Creswell (2018) mentioned that it is a mix of both qualitative and quantitative data in a study. In this mixture, the qualitative data tends to be open-ended without predetermined responses, while quantitative data is in charge of closed-ended responses. Moreover, Sharma, Bidari, Bidari, Neupane and Sapkota (2023) said that the use of this type of research enables researchers to answer research questions with adequate depth and width, also, it assists to generalize findings and implications of the researched issues to a whole population of the field; additionally, when this mixed approach is used provides a better understanding of the research problem, the implementación of both types of data sets allow the research question in a way that its answer is greater certainly and wider implications to the final answer of the research question creating, in this way, a complete and rigorous answer.

The quantitative features can be found in the raw counts and descriptive statistics created to illustrate the most common non-target-like use learners tend to use in oral speech

when they use the nominal or verbal inflection, and the qualitative aspect is related to the relationship between those numbers with some of the possible criteria of accuracy or emergence as stated in the literature review.

Research Questions

1. What inflectional patterns appear in the oral production of Mexican learners of English?
2. What is the order of inflectional patterns observed in the oral production of English as foreign language learners?

General objective:

To analyze nominal and verbal inflection patterns in the oral vocabulary production of learners of English as a foreign language.

Specific Objectives

1. To identify the nominal and verbal inflection patterns in learners' oral production.
2. To describe the nominal and verbal inflection patterns in learners' oral production according to established emergence and accuracy criteria.

The Corpus

The data analyzed consisted of the transcripts of oral interviews previously collected and transcribed by the MexLeC research team. As described earlier in Chapter 2, these interviews are composed of four distinct tasks, each designed to elicit different types of language production, ranging from personal descriptions and narratives to expressions of preferences and opinions. This structured design allows for the examination of various aspects of learners' linguistic performance across different communicative contexts.

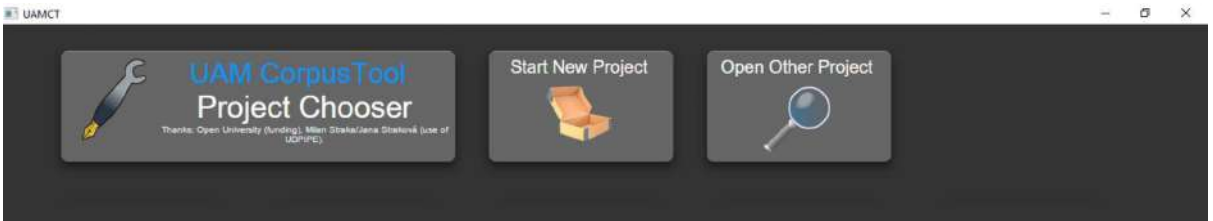
The present study specifically focuses on exploring how vocabulary patterns are learned and used by Mexican learners of English as a foreign language. To this end, the cohort 001B-001 from the MexLeC corpus was selected for analysis. This particular cohort was chosen because the university to which it belongs admits students without requiring a specific level of English proficiency at entry. Consequently, it is expected that these participants display a greater range of language development, including a higher frequency of lexical errors and variation in vocabulary use. Analyzing this group provides valuable insight into the early stages of lexical acquisition in foreign language contexts and contributes to understanding how learners with limited prior exposure to English begin to build and organize their vocabulary system.

Data analysis instrument

Universidad Autónoma de Madrid (UAM) Corpus Tool is a software designed for analyzing the morphological features of a corpus O'Donnell (2008), it enables users to annotate a corpus of text files at various linguistic layers, which can be customized by the user. Users can create a hierarchy of tags for each layer using a graphical interface and then annotate the text by selecting segments and assigning features from the tag hierarchy. Additionally, this free software can be downloaded at no cost from its website. Once a corpus project is created, users can perform a wide range of functions, including creating tagging schemes and conducting statistical analysis of the data, a screenshot of the home page is shown in figure 1. For the purpose of this research, inflectional tagsets were obtained from Ardébol (2012), Fontana (2013), and Demarta (2012) and the selected corpus texts were tagged in accordance. The detailed procedure and some relevant examples are described in the stages of the research section.

Figure 1

UAM Corpus Tool aspect before opening a project



Stages of the research process

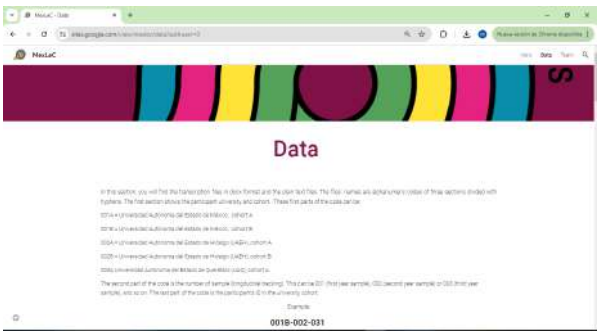
I. Selection and download of transcripts.

First of all, it was essential to begin by selecting the appropriate interviews from the Mexican Learner Corpus (MexLeC), which is a valuable resource that contains a wide range of recorded and transcribed interviews with English learned from different universities in Mexico. This diversity allows for a more representative and comprehensive analysis of English language learners in the Mexican context.

The interviews were carefully selected from MexLeC based on the most basic level of English language proficiency that corresponds to the first year of data collection, as it can be seen in Figure 2. This initial stage of language learning was chosen to allow for a focused analysis of beginner-level English usage among Mexican learners. Once all the interviews from the corpus were identified, the next step involved downloading the corresponding files in preparation for the cleaning process. After downloading the documents, they were saved in .docx format to facilitate easier editing and formatting. The cleaning process required the removal of all non-verbal annotations and transcription symbols found between brackets; these included tags as <name>, <er>, <laugh>, <xx>, <address>, <place>, <.>, <...>, and <foreign=>. Eliminating these elements was essential in order to produce a cleaner and more coherent version of the transcripts that would allow for more accurate linguistic analysis.

Figure 2

Home page screenshot of MexLeC corpus website

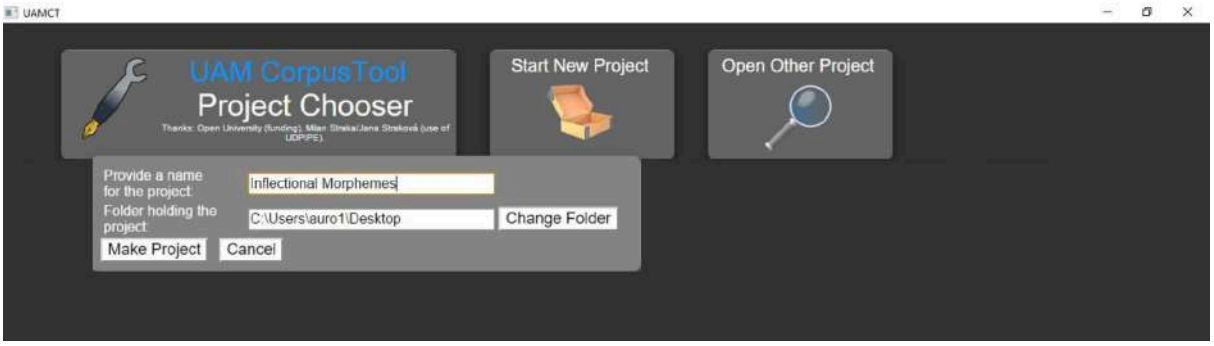


II. Use of UAM Corpus Tool

The second step was to create a new project in UAM CorpusTool, which I named "Inflectional Morphemes", as can be observed in Figure 3 below. This project served as the central space for organizing and managing all the data and annotations related to the linguistic analysis. Once the project was set up, the following step involved creating a set of specific tags that corresponded to the morphological features that were going to be examined throughout the study. These tags were designed to capture the key inflectional morphemes that are present in the learner's spoken English in this study. The analysis focused on nominal and verbal inflection. For nominal inflection, the tags used were for regular and irregular plurals. On the other hand, for verbal inflection, the tags used were irregular and regular past tense forms, the third person singular present tense, and the progressive aspect. Additionally, the analysis included tagging of lexical verbs and auxiliary forms of the verb *to be* to better understand how these elements were used by learners at a basic proficiency level. Defining and applying these tags was essential to ensure consistency in the analysis to enable meaningful observations regarding the development and use of inflectional morphemes in learners' language.

Figure 3

Creation of a new project

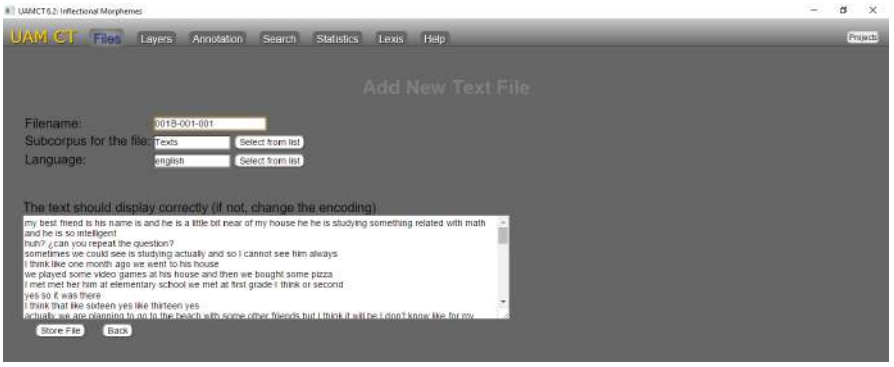


III. Loading the data

Once the interviews were successfully downloaded and carefully cleaned, the next step was to ensure that only the relevant spoken content remained in each transcript. This cleaning process involved systematically removing various non-essential elements, such as long pauses, speaker turns, interjections, inaudible parts, extralinguistic information, incomplete words of false starts, and foreign language words that were not part of English utterances. The cleaned transcripts were copied and pasted into UAM CorpusTool as shown in Figure 4.

Figure 4

Emptying of dataset



Each file was then saved with a clear and descriptive name, which is formed with numbers as it was done on the official website; this code reflects the university and cohort of collection, as well as the year of collection (first year, second or third) and the ID of a given participant, so that each file can be easily located and referenced during later stages of analysis. This step was particularly important in cases where it might be necessary to return to an individual interview for closed examination or comparison.

IV. Tagging and data analysis

Finally, after completing all the previous stages, the next stage involved the creation of the tags that would be used to analyze all the selected files from the corpus systematically. These tags were essential for identifying and categorizing the specific linguistic features relevant to the objectives of this research.

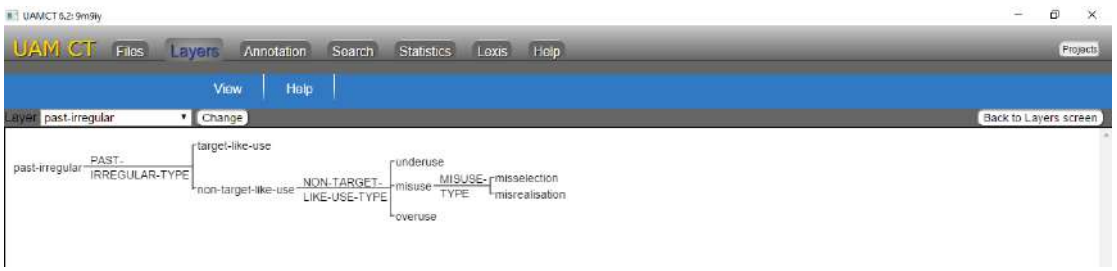
For the purpose of this study, the focus was placed on the analysis of inflectional morphemes. The selected elements included two noun-related categories and six verb-related categories. The noun-related categories focused on the use of the regular and irregular plural forms, while the verb-related categories included the regular past tense, irregular past tense, third person singular present, progressive-ing, be lexical, and be auxiliary of the verb *to be,* for example, figure 5 presents how layers were done. These categories were chosen based on their significance in early stages of English language learning and their relevance in tracking learners' grammatical development.

After the morphemes to be analyzed were clearly defined, a total of seven annotation layers were created within the UAM CorpusTool titled "Morphological Inflection". Each of these seven layers corresponded to one of the targeted inflectional morphemes and was structured with a consistent set of tags, ensuring uniformity across all layers. This consistency was critical for generating reliable statistics and frequency counts, which would later be used to answer the research question.

To illustrate the structure used in the annotation process, the following figure presents the layout of the "Past Irregular" layer with the UAM Corpus Tool project. The structure applied to each morpheme layer followed the same model, allowing for systematic annotation and analysis of each morphological feature across the dataset.

Figure 5

Past irregular Tag- set sample



To end with the description of the tagging procedure, some excerpts from the corpora are presented below, together with their corresponding categorisation tags and brief explanations:

Non-target-like use: This category includes instances where the morpheme was either used incorrectly or in contexts where it was not obligatory (SNOC), as in the following sentence:

(1) "I think his mother is appear on the background."

As we can observe in the example, the verb requires its "ing" form to accurately form the participle. This tag is divided into three main categories: underuse and misuse (which is subdivided into misselection and misrealization), and overuse.

Underuse is a subcategory of non-target-like use that covers cases where the morpheme was not used despite being required, as in the following example:

(2) " *I was just think like English and French but when in the first day…"*

In this example, the progressive form of the verb (i.e., thinking) is required, but it was not used despite being mandatory.

Misuse is divided into two branches that are misselection and misrealization, accounting for instances where the student used a morpheme but did so incorrectly. Misselection is used for cases where the student selected a grammatically correct morpheme but used it in the wrong context, meaning the chosen morpheme was not the one required, as in the following sentence:

(3) *"...his father is trying to make the car works but in the third picture I can see…"*

As it can be observed, the word *works* in non-target-like use because even though the sentence is grammatically correct, it is used in the wrong context due to the nature of the task, the context it is needed is in the past. On the other hand, misrealization is for instances where students supplied a form that does not exist in English instead of using the required

morpheme. In this case, during the analysis, no morphemes were found in this subcategory, but some examples are lexemes as *singed*, *breaked*, and so on; basically, they are words that do not exist in the English language.

Finally, the tag overuse, corresponds directly to SNOC (suppliance in non-obligatory contexts), referring to instances where a morpheme was used inappropriately for the given context as in the following example:

(4) "I decide my family because they are the persons who impulse you…"

In this case, the word *persons* is part of the non-target-like use due to the overgeneralization of the grammatical rule where the correct morpheme for the context would be the word *people*.

The final step was to calculate the emergence and accuracy criteria. The first step was simply calculated by computing the number of tokens, for each inflectional category, appearing in the corpus, considering a rate of four tokens and above. To compute accuracy level it was computed the TLU (Target Language Use) formula, applied in Fontana (2013):

TLU= SOC/ OC+SNOC

Where=

SOC is Suppliance in Obligatory Context;

OC means Obligatory Context

SNOC means Suppliance in Non-Obligatory Context

The findings derived from applying the previously outlined procedures to the entire set of materials in the MexLeC corpus presented in the following chapter.

Chapter 4. Results

In this chapter, the results of the study are presented and analyzed through the criteria of accuracy and emergence. For this reason, this chapter is divided into three sections: the first one presents the general results regarding the target-like use, non-target-like use and overuse to give a general idea of them. In the second section, the results are presented according to the criteria of accuracy, and finally, in the third section the results are presented according to emergence criteria.

4.1 General Results

In relation to the general results, tokens were counted of each one of the inflectional paradigms which are organized from the most frequent to the least frequent in terms of the usage of the morphemes in the twenty four interviews from MexLeC.

In order to obtain the general frequencies (i.e., the Obligatory Contexts, OC) of each morpheme, the results of target-like use and non-target-like use were summed since both were taken into consideration. Consequently, the percentage of overuse could exceed 100% when calculating the relative frequencies. The overuse percentage thus would result in more than 100% for the relative frequencies.

The general frequencies of the tokens of target-like use, non-target-like use, and overuse in each of the inflectional patterns studied of *regular and irregular plural*, *lexical be*, *auxiliary be, regular and irregular past*, t*hird person of the singular*, and *progressive* are included in Table 1 below.

Table 1

General results of inflectional morphemes

Inflectional type	Target-like use		Non-Target-like use		Overuse		Obligatory Contexts
	Frequency	%	Frequency	%	Frequency	%	
Regular Plural	304	96,82	10	3,18	6	1,91	314
Lexical Be	257	96,62	9	3,38	0	0,00	266
Irregular Past	106	81,54	24	18,46	2	1,54	130
Regular Past	65	81,25	15	18,75	0	0,00	80
3rd. singular	55	78,57	15	21,43	1	1,43	70
Auxiliary Be	40	86,96	6	13,04	2	4,35	46
Progressive	36	85,71	6	14,29	1	2,38	42
Irregular Plural	17	80,95	4	19,05	4	19,05	21

It can be observed that there are different frequencies of the morphological morphemes. The inflectional morpheme with the highest frequency of use during the interviews is *regular plural* with 320 tokens in total, while 304 (96.825) tokens are target-like use, ten (3.18%) are non-target-like use, and six (1.91%) tokens correspond to overuse.

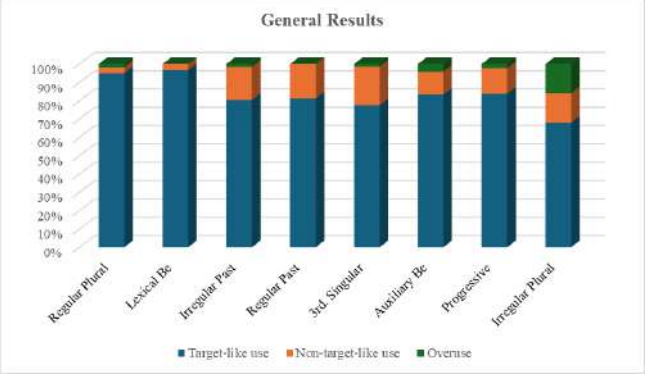
Lexical be is the second most frequent morpheme with a total of 266 tokens. It shows a very high accuracy rate with 257 (96.62%) target-like use tokens, and only nine (3.38%) instances of non-target-like use, and no overuse was identified. Then, *irregular past* contains

130 tokens, with 106 (81.54%) reflecting target-like use, 24 (18.46%) showing non-target-like use, and two (1.54%) containing overuse tags. Meanwhile, *regular past* contains 80 obligatory context tokens; this category has 50 fewer tokens in comparison with past irregular. It shows that 65 (81.25%) reflect in target-like use, 15 (18.75%) non-target-like use and zero instances of overuse.

The *third singular* contains 55 tokens (78.57%) of target-like use, 15 (21.43%) tokens of non-target-like use of a total of 70 tokens, plus one (1.43%) token that belongs to overuse. In addition, *the auxiliary be* has a total of 46 tokens where 40 (86.96%) tokens show target-like and six (13.04%) tokens non-target-like use and two (4.35%) tokens reflect overuse. *Progressive* contains 42 tokens where 36 (85.71%) tokens belong to target-like use and six (14.29%) to non-target-like use and only one (2.38%) token belongs to overuse tag. Finally, the morpheme with the least usage is the *plural irregular* with 21 tokens in which 17 (80.95%) tokens reflect target-like use, four (19.05%) non-target-like use, and four (19.05%) are used in an overuse manner. In this last case, it was observable that the regular plural is the most used morpheme with a total of 314 tokens; nevertheless, the irregular plural was the least morpheme used by learners with a total of 25 tokens. A visual representation of results related to target-like use, non-target-like use, and overuse can be observed in Figure 6.

Figure 6

Graphic of general results with target-like use, non-target-like use, and overuse



As Figure 6 shows, the frequency and accuracy of morphological morpheme use varies among learners. The *regular plural* morpheme, with the highest frequency of 320 tokens, is notably acquired, demonstrating a high target-like accuracy of 96.82% and minimal non-target-like and overuse instances. Similarly, the *lexical be*, ranking second in frequency, shows strong mastery, with a target-like use of 96.62% and no overuse was observed. In contrast, morphemes such as the *regular past*, *third-person singular*, and *irregular plural* display significantly lower frequencies and accuracy rates. The *irregular plural* morpheme, with the lowest usage frequency (21 tokens), highlights a gap in consistent use, with considerable non-target-like and overuse, emphasizing its challenges for learners.
This analysis highlights the need for targeted reinforcement of less frequent and more challenging morphemes, like the *irregular plural*, which shows the largest discrepancy in usage compared to high-frequency morphemes such as the regular plural.

4.3 Emergence Criterion Results

As it was explained before in the section 1.4.1 Emergence Criterion, is an aspect of language learning indicating whether a morpheme of structure emerged in the learner schema. As it was previously mentioned the criterion to determine if an inflectional morpheme has emerged is if the word appears accurately at least four times in the students' speech. In the present study, the analysis focuses on determining whether specific functors have emerged across the entire Cohort B. It is important to note that the participants are A2-level learners of English, a factor that is taken into account in order to ensure a more objective interpretation of the results. Results on the emergence criteria are shown in table 2.

Table 2

Emergence criterion results

Inflectional Type	Frequency	Emergence Criterion (>4)
Lexical Be	266	Yes
3rd. singular	71	Yes
Regular Plural	320	Yes
Progressive	43	Yes
Irregular Past	132	Yes
Irregular Plural	25	Yes
Regular Past	80	Yes
Auxiliary Be	48	Yes

As it can be observed in Table 2, lexical *be* is used 266 times, while the *third-person singular* form occurs 71 times. Regarding number, regular plural forms occur 320 times, while irregular plural forms appear 25 times. In terms of tense, *irregular past* forms contain a total of 132 tokens, whereas regular past forms contain 80 tokens. The *progressive* inflection is recorded 43 times, and the *auxiliary be* appears in 48 instances.

These results provide evidence that learners from Cohort B of the MexLeC Corpus have acquired these morphemes since each morpheme obtained counts of at least four tokens according to the emergence criterion established by Palloti (2017).

All these results, displayed in Table 2 and supported by literature, provide evidence for the acquisition of all the analyzed morphemes, as each one obtained more than four tokens. For this reason we can claim that cohort B from the MexLeC corpus, inflectional morphemes have been acquired according to emergence criterion.

4.3 Accuracy Criterion Results

In order to analyze the accuracy results, Brown's formula Suppliance in Obligatory Context formula (SOC, as cited in Fontana, 2013) was used. According to this formula, accuracy rate, is calculated by dividing the frequency of Target-Like Use (TLU) by the Obligatory Context (OC, i.e, TLU+NTLU), plus the Suppliance in Non-Obligatory Context

(SNOC, i.e., overuse), whose result must be at least 0.90 to determine whether a morpheme or structure has been acquired by accuracy criterion. Table 3 presents the accuracy rate of each morpheme, arranged in rank order.

Table 3

Accuracy rate

Inflectional type	Suppliance in Obligatory Context **(SOC)**	Obligatory Context **(OC)**	Suppliance in Non- Obligatory Context **(SNOC)**	OC+SNOC	Accuracy rate
Lexical Be	257	266	0	266	0,97
Regular Plural	304	314	6	320	0,95
Progressive	36	42	1	43	0,84
Auxiliary Be	40	46	2	48	0,83
Past regular	65	80	0	80	0,81
Past irregular	106	130	2	132	0,80
3rd. singular	55	70	1	71	0,77
Irregular Plural	17	21	4	25	0,68

As it can be observed in Table 3, the inflectional morpheme with the highest accuracy rate is *lexical be* with 0.97, followed by the *regular plural* with 0.95, showing that there are just 0.02 of differences in the range of accuracy. According to Brown (1973), an accuracy rate of at least 0.90 is needed to be able to claim that the morpheme is accurately used. Therefore, it can be argued that both *lexical be* and the *regular plural* have been used accurately. By saying that, it can be established that the morphemes that have been used the an accurately way are lexical be and plural regular.

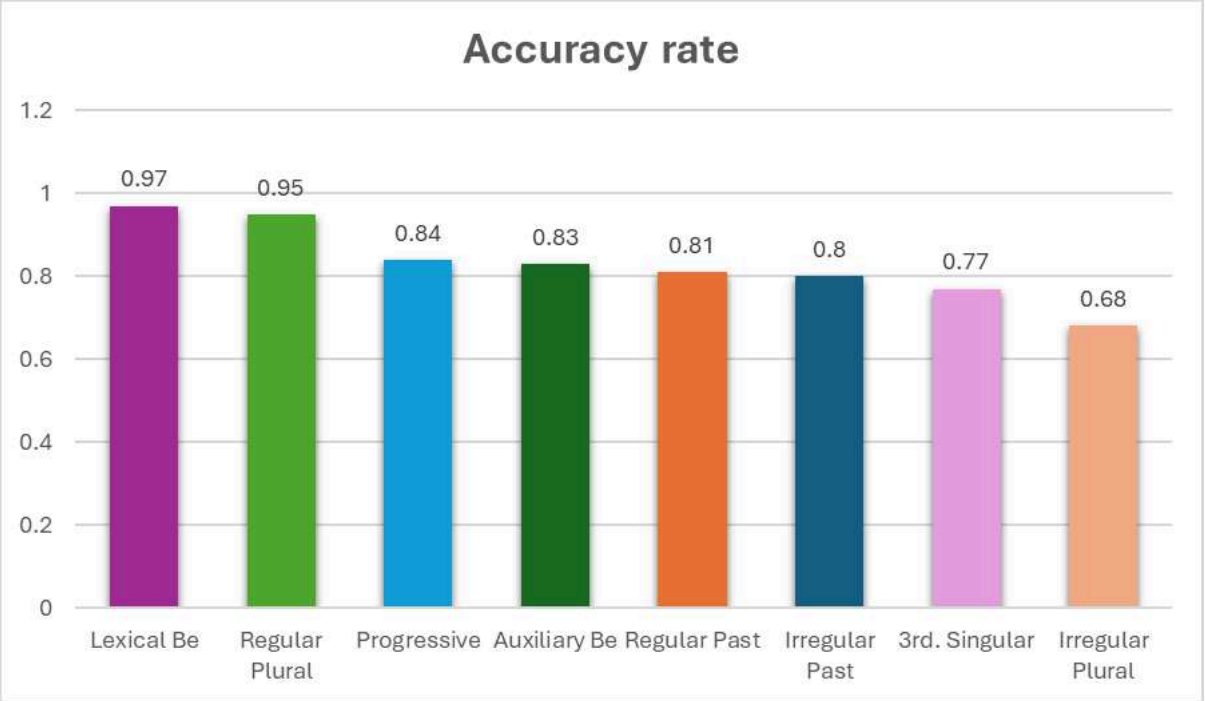
On the other hand, the other morphemes have not reached the accuracy rate required: *progressive*, *auxiliary be*, *regular past*, *third person singular*, and *irregular plural*. The *progressive* obtained an accuracy rate of 0.84 which shows a significant decrease of 0.11 in comparison with the *regular plural*. Then, the *auxiliary be* has an accuracy rate of 0.83, and the *regular past* verb has 0.81 which shows a slight difference of 0.01 with the *irregular past* with 0.80 level of accuracy.

Nevertheless, the third *person singular* has an accuracy rate of 0.77, while the *irregular plural* has 0.68, which shows that these two morphemes are the ones that learners need to work more on. These results indicate that these two morphemes present greater difficulty for learners and therefore require further attention.

Finally, to better illustrate the results and allow for observation of the emergence pattern, Figure 7 presents the accuracy rate in order of all the morphemes studied in this research; lexical be (0.97), plural regular (0.95), progressive (0.84), auxiliary be (0.83), past regular (0.81), past irregular (0.80), third person of the singular (0.77), and plural irregular (0.68).

Figure 7

Accuracy rate graphic



In summary, the data revealed that the morphemes with the highest levels of accuracy, specifically lexical be (0.97) and regular plural (0.95), meet the accuracy criterion of 0.90 established by Brown (1973), signifying strong, reliable acquisition of these forms. These morphemes are used accurately and consistently by learners, with only a slight difference in accuracy rates between them. However, several other morphemes, including the progressive, auxiliary be, regular and irregular past tense, third-person singular, and irregular plural, have not yet reached this 0.90 criterion, indicating that they are still in the process of being learned. Notably, progressive (0.84) and auxiliary be (0.83) are approaching high accuracy, while forms like third-person singular (0.77) and irregular plural (0.68) display the lowest accuracy, highlighting areas that may benefit from further targeted learning interventions.

Chapter 5. Discussion

To provide a better understanding of this study, it is important to remember our general aim, which is to analyze the nominal and verbal inflectional patterns in learners' oral production in the acquisition of English as a Foreign Language. Furthermore, specific objectives are the following: to identify the nominal and verbal inflectional patterns in learners in the acquisition of English as a foreign language, to describe the nominal and verbal inflection patterns in learners' oral production according to established emergence and accuracy criteria.

Patterns of nominal and verbal inflection in learners' oral production

As stated in Chapter 1, inflection is the word process in which a word changes its form so it can fit into different contexts without changing the grammatical category. Nominal inflection refers to the change that a noun could have, and the information that can be obtained from that inflection is number, case, gender, and noun class. On the other hand, verb inflection is the change that verbs have in different contexts, and the information that can be gathered from that verbal inflection is person, tense and aspect, mood and modality, and voice. All these features are important to transmit a complete message with that implicit information.

In the case of nominal inflection in English, the plural patterns are regular and irregular, where the regular plural marking is the suffix -s, such as in boy**s**, girl**s** or house**s**. However, there are some nouns that are formed irregularly, which means that they can change their form completely or may not have clear patterns like in mouse-mice or man-men.

Verbal inflection is wider and can be found in verbs as the *lexical* and *auxiliary be*, *regular* and *irregular past*, *progressive*, and *third person of the singular. Lexical be* is in charge of linking the subject of the clause to the following word that describes, identifies, or

locates the subject, while the *auxiliary be* is used together with a main verb to show the tense of the verb. In the *third person of the singular*, it is present the marking -s to express that it is a singular subject and it is in the present simple tense. However, there are also regular and irregular past in which the regular is identified by the -ed in verbs that are regular; nevertheless, irregular, usually changing their form, or they do not. Finally, the *progressive*, which frequently expresses ongoing action, is formed by the auxiliary be plus the present participle.

Patterns of nominal and verbal inflection considering emergence criteria in learners' oral production.

As it was mentioned in section 1.3.1 emergence criterion is the first systematic use of a structure at a point in time, which can be identified when learners have understood the learning task; additionally, the use of four tokens are needed to indicate that the morpheme has emerged. The results showed that all the inflectional morphemes have emerged; however, it cannot be claimed the way in which they were emerging due to the fact that this is not a longitudinal study.

All inflectional patterns analyzed met the established emergence criterion, as each was produced at least four times in the learner's speech. This indicates that forms such as regular and irregular plural, regular and irregular past, third person singular, lexical be, auxiliary be, and the progressive have emerged and can be considered part of the learners' developing linguistic system. Although not all productions were consistently accurate, the results suggest that learners have begun to integrate these inflectional patterns into their linguistic repertoire.

It is important to note that, unlike Zhang and Widyastuti (2010), who examined emergence through less controlled data, the present study elicited language through a

structured interview designed to prompt the use of specific grammatical forms. This methodological difference may account for the relatively high emergence rates observed. In this way, even though learners did not produce all the inflectional patterns correctly, they could integrate these patterns in their linguistic repertoire.

Certain patterns emerged with more frequency that go from the simpler forms to the more complex ones (Zhang & Widyastuti, 2010). For instance, I found that the patterns with more frequency are *progressive*, *regular past*, *lexical be*, *third person singular, regular plural*, *auxiliary be*, *irregular plural*, and *irregular past*. It can be implied that these patterns are becoming automatic and without conscious effort. In contrast to Zhang and Widyastuti who claimed that regular plural did not emerge in any of their three informants, I found in this study that this morpheme has already emerged in the group taken as a sample. However, the progressive and plural forms have emerged in both studies.

These inflectional patterns exhibited the required emergence rate provided evidence that they have become an integral part of the learners' language use. Beyond correct usage, the patterns are applied in different linguistic contexts, signaling principally an understanding of inflectional morphology. The ability to transfer these patterns during speaking is a hallmark of acquisition, so it demonstrates that they are already internalized knowledge.

Patterns of nominal and verbal inflection considering accuracy criteria in learners' oral production.

As it was mentioned in section 1.3.2 Accuracy Criterion, accuracy is the production of speech without errors. For this study, accuracy was computed using the target language use (TLU) formula, considering an accuracy rate of 0.90 as the criterion to determine whether a morpheme has been fully learned by the learners. This limit of 90% accuracy rate ensures that the morpheme is not only familiar to the learners, but it is also used correctly with high

consistency across different contexts (Brown, 1973). A morpheme with this percentage or higher is considered learned as it reflects a strong and reliable understanding of its form and application in the different linguistic contexts. Applying this criterion of accuracy was helpful to differentiate between morphemes that are already integrated into the learners' language system and those that are still in development.

Regarding the results, some forms stand out for their precision in usage. For instance, the *lexical be* morpheme presented an accuracy rate of 0.97 in this way, it indicates that learners are using the *lexical be* correctly in nearly all instances. These results are aligned with Fontana (2013) and Akbaş and Dinçer (2021), who reported a score of 0.92 for *lexical be*, making it one of the highest morphemes with target- like use.This high score suggests that *lexical be* is solidly learned and is likely applied almost automatically in various contexts.

Similarly, the *regular plural* morpheme obtained a high score, with an accuracy rate of 0.95, which shows learners' understanding of the rules of regular pluralization and their ability to consistently produce this form accurately. This high score suggests that learners have largely internalized the rule of adding the plural -s to nouns in obligatory contexts. This result is consistent with Altarawneh and Hajjo's (2018) findings, who claimed that learners show awareness of regular plural nouns and are generally able to recognize and use them correctly.

The *progressive* morpheme, although slightly below the 0.90 threshold with an accuracy rate of 0.84, still shows a relatively high level of accuracy. This result is consistent with the findings of Demarta (2012), who reported a similar percentage of 84%. Although this morpheme does not yet meet the strict acquisition criterion, the accuracy rate suggests that learners are developing a good command of the progressive form and are able to use it correctly in most contexts.

On the other hand, morphemes that fall into the medium accuracy rate are auxiliary *be* and *regular past* which provide insights into forms that learners are generally able to use correctly but have not yet fully mastered to the point of consistent accuracy across all linguistic contexts. These medium-range scores suggest that learners are developing proficiency in the enforcement of these morphemes, though some inconsistency remains. For example, the auxiliary *be* obtained an accuracy rate of 0.83, indicating that learners are adept at using this form in most situations. Similar results were found by Fontana (2013) with 89.02% and Demarta (2012) 79.57%, respectively. However, the presence of occasional errors implies that they may not have completely internalized the specific rules governing its use, especially in more complex or less familiar contexts. This pattern of partial acquisition suggests that while the *auxiliary be* is generally understood, learners may still be refining their use in the contextual shifts that influence its correct application.

In a similar way, the *regular past* morpheme showed an accuracy rate of 81%, reflecting that learners have an understanding of forming the regular past tense but may still make occasional errors in its use. In this sense, Akbaş and Dinçer (2021) reported a similar result of 80%; this supports the idea that learners tend to acquire the regular past form with a medium level of accuracy as their proficiency develops. These scores indicate that while learners are confident using the regular past structure and generally apply it correctly, they may practice the basic rule of regular past by adding "-ed" to form the past tense in regular verbs in spontaneous language production without lapsing into inaccuracies. These medium accuracy scores highlight morphemes that are within reach of full acquisition but require further reinforcement for learners to achieve full mastery and automaticity in usage.

With regards to the category of lower accuracy scores, I found a group of morphemes that learners appear to struggle with more frequently and for which consistent, accurate usage has not been achieved yet. These morphemes represent the more challenging aspects of the

learners' language development process and indicate areas where irregular and exception-based forms may be hindering full acquisition. Specifically, morphemes such as the *irregular past* form, which has an accuracy rate of 0.80, present learners with some challenges.

This relatively lower accuracy suggests that while learners may be somewhat familiar with the past irregular forms, their use is not fully consistent or reliable yet. The difficulty likely arises from the fact that irregular past forms do not follow a predictable pattern, which requires learners to memorize individual forms rather than rely on a standard rule. As a result, learners may struggle to recall or apply the correct form in spontaneous language production, leading to a notable rate of error.

Another morpheme with a low accuracy rate is the *third person singular form* with 0.77. This number is higher in comparison with Fontana (2018); Ardebol (2012) and Demarta (2012) who got results lower than expected, going from 15.28%, 17.22% and 18.18% respectively. These scores indicate that while some learners may master the basic rule of adding "-s" or "-es" to verbs in third-person singular contexts, there is still a considerable margin of error. Even though these percentages are lower than the one obtained in this research, it is important to mention that they belong to the lowest categories from the studies. The challenge with this morpheme may stem from the fact that *third-person singular forms* require learners to keep track of the relationship between the subject and verb agreement, an area where errors are common due to the need to constantly monitor and apply this rule based on the subject's number and person. Additionally, the rules governing the third person singular can vary slightly depending on the verb ending, as "-ed" for verbs ending in certain consonants, adding another layer of complexity for learners.

The *irregular plural* form, with an accuracy of only 0.68, represents one of the most difficult morphemes for learners to master in this category of low accuracy scores. According

to Altarawneh and Hajjo (2018) this low score highlights the significant challenges posed by irregular plural forms, which often require more memorization of unique forms rather than the simple addition of "-s" or "-es"; so, the overgeneralization phenomenon is presented. For instance, learners may struggle to remember that "child" becomes "children" rather than "childrens". The lack of a consistent rule for forming these irregular plurals means that learners must rely heavily on memory, and without frequent reinforcement or exposure, these forms are more prone to error and misuse (Khor, 2012).

The classification of morphemes into high, medium, and low accuracy scores offers a clear and structured overview of learners' language acquisition process, illustrating which morphemes have been fully mastered, which are still developing, and which continue to pose challenges. By identifying morphemes that fall into the low accuracy range, such as *irregular past forms*, *third person singular*, and *irregular plurals*, it can help to better understand the specific areas that learners require additional support and practice. These lower accuracy scores generate an impact of irregularity and exceptions on language learning, emphasizing the need for sustained exposure and practice to help learners overcome the obstacles posed by these complex forms (Altarawneh & Hajjo, 2018).

Focusing on nominal inflection, there was a notable distinction between the accuracy rates of *regular* and *irregular* plural morphemes, which demonstrates a clear developmental sequence in the acquisition of these forms. The first morpheme that learners consistently acquire with high accuracy is the regular plural, showing a great accuracy rate of 0.95. This may happen due to the simplicity and consistency of their rules that make regular plural nouns easier for learners to internalize and apply accurately across diverse contexts. In contrast, the last nominal morpheme to be learned is the *irregular plural*, with only 0.68, which has a significantly low score, pointing to the inherent difficulty *of irregular plural* forms that do not follow the predictable pattern seen in regular plurals. The accuracy

difference between regular and irregular plurals is 0.27, which illustrates the challenges learners face when encountering morphemes that require memorization rather than the application of a straightforward rule.

When examining the sequence of learning for verbal inflections through the accuracy rates, I found that there was a certain difficulty associated with each morpheme. At the top of the list, with a high accuracy rate of 0.97 is the *lexical be* which is one of the morphemes used more frequently in everyday language in its essential role in constructing sentences. Following the *lexical be* is the *progressive* form, with an accuracy rate of 0.84, this morpheme may be relatively easier for learners to acquire because it follows a predictable pattern and it is frequently used in common expressions and descriptions of actions in comparison with the accuracy score of *lexical be*. These results showed that learners are fairly confident in forming the progressive but they may still have some minor inconsistencies.

Next in the accuracy order is the *auxiliary be*, with an accuracy score of 0.83. Although this auxiliary form is closely related to *lexical be*, it requires learners to apply it in a slightly more complex grammatical structure, as it is used to support other verbs rather than act as the main verb in a sentence. The moderate accuracy score for the *auxiliary be* suggests that learners are familiar with this form but may encounter occasional difficulties in its application.

The *regular past* tense comes next with an accuracy rate of 0.81. The difference between the previous accuracy scores with the last forms that are mentioned below as irregular past and *third person singular* could indicate that while learners understand the basic rule for regular past tense formation, they may still experience occasional lapses in spontaneous language production or when forming past tense with less common verbs. In contrast, the *irregular past* form, scoring an accuracy rate of 0.80, comes immediately after

the regular past tense. The slightly lower accuracy for the *irregular past* tense suggests that learners find this form more challenging due to the lack of a consistent rule, which requires individual memorization of each irregular verb form.

Finally, at the lowest end of the accuracy order of the verbal inflection is the *third-person singular* form, with an accuracy rate of 0.77. This score suggests that while learners may be aware of the rule to add "-s" or "-es" to verbs in the third person singular present tense, they struggle with its consistent application. The third-person singular morpheme may be challenging because it requires constant attention to subject-verb agreement, an aspect of grammar that can be easily overlooked (Haspelmath,2002). This lower accuracy rate highlights that learners may need additional practice and reinforcement with the third-person singular.

The analysis of accuracy rates across the different morphemes reveals the general order of learning accurately in both nominal and verbal inflection; firstly, the high score accuracy rates are *lexical be*, *plural regular*, and *progressive*, in which learners demonstrated a full understanding of the usage of those morphemes. Followed by the medium score accuracy rates that are *auxiliary be* and *past regular*, where learners may have an understanding of the usages of the morphemes, but there are some inconsistencies at the moment of applying them in different linguistic contexts. Finally, the low score accuracy rates are *past irregular*, *third person singular* and *plural irregular*, where learners may be aware of the specifications in the use of each morpheme, but they are not really related to these morphemes.

Chapter 6. Conclusions

As stated in the introduction, the aim of this research is to analyze patterns of nominal and verbal inflection in vocabulary learning in learners' oral production of English as a foreign language. The relevance of this study relies on the lack of corpus-based research that explores the specific linguistic errors made by Mexican speakers studying a degree in English Language Teaching.

This analysis of nominal and verbal inflection patterns in the vocabulary learning of learners' oral production in English as a Foreign Language highlights key morphological structures that are crucial for effective oral communication in this study, eight inflectional morphemes were identified: *plural regular* and *irregular*, *lexical* and *auxiliary be*, *regular* and *irregular past*, *progressive*, and *third person of singular*. These findings answer the first research question by showing the nominal and verbal inflection patterns that appear in learners' oral production and fulfill the objective of identifying these patterns.

The identification of plural forms, both regular and irregular, in nominal inflection make emphasis on the complexity learners face when forming nouns in the target language. Similarly, verbal inflections such as the *lexical* and *auxiliary be*, the *regular* and *irregular past* tense forms, the *third-person singular*, and the *progressive* aspect demonstrate how learners use tense, aspect, mood, and modality, case, and voice. These inflection patterns are fundamental to expressing time, person, number, and ongoing actions, and their learning plays a significant role in the learners' ability to produce grammatically accurate and contextually appropriate speech.

On the other hand, morphemes like the *irregular past*, *third-person singular*, and *irregular plural* exhibit lower accuracy rates, highlighting areas where learners face greater challenges. These irregular forms often lack predictable patterns, making them more difficult

for learners to master. The analysis of these lower accuracy scores reveals the need for further practice and reinforcement, especially with forms that require memorization rather than rule application.

In both nominal and verbal inflection, the sequence of learning follows a logical pattern, with regular and more predictable forms being learned first, followed by more complex and irregular forms. The data suggest that while learners are generally able to use many of the morphemes correctly, continued exposure and targeted practice are necessary for achieving full mastery, especially in the case of irregular forms that require more cognitive effort and attention.

Regarding the order of inflectional patterns observed in the oral production of English as foreign language learners, the study showed that the sequence typically begins with the learning of the lexical be, followed by regular plurals and the progressive form. Afterward, learners tend to acquire the auxiliary be, then move on to regular past tense forms, followed by irregular past tense forms. The learning process continues with the third-person singular inflection, and finally, learners master irregular plural forms. This finding confirms that the order patterns identified in this study follows a progression from regular and frequent forms to more complex and irregular ones.

One of the limitations of this research was the relatively small number of interviews conducted from the selected corpus. Since the group of participants was carefully chosen, the sample size was limited, which may have restricted the diversity of linguistic data collected. A larger sample size could have provided a more extensive range of cases, thereby offering a deeper and more comprehensive understanding of the learning of inflectional morphemes in both nominal and verbal inflection. By including a wider variety of participants, the study could have better captured the nuances and variability in the use of these morphological features, potentially leading to more robust conclusions about the patterns and processes

involved in their acquisition. This would have strengthened the findings and contributed to a more detailed analysis of the linguistic phenomena under investigation.

Future research

For future research, it would be beneficial to conduct a longitudinal study using the same group of participants analyzed in this study. By following the same individuals over an extended period, researchers could observe how their English language skills develop and evolve, particularly focusing on the learning of inflectional morphemes. This approach would provide valuable insights into the learners' progress, allowing for a detailed examination of the stages and patterns in their mastery of both nominal and verbal inflections over time.

Additionally, expanding the scope of the study to include the learning of inflection in adjectives could offer a more comprehensive understanding of morphological development. Tracking learners from the beginner level through to more advanced stages would enable researchers to observe how they gradually acquire and correctly apply adjectival inflections, such as comparative and superlative forms. This expanded focus could reveal interesting developmental trajectories and challenges that learners face as they increase their proficiency, shedding light on common errors and the mechanisms behind their improvement. By examining these aspects in a longitudinal framework, future research could contribute valuable data on how different morphological features are learned across various stages of language learning, ultimately providing a more holistic picture of the linguistic development process.

References

Akbaş, E., & Dinçer, Z. (2021). Accuracy order in L2 grammatical morphemes: Corpus evidence from different proficiency levels of Turkish learners of English. *Studies in Second Language Learning and Teaching, 11*(4), 607–627. https://files.eric.ed.gov/fulltext/EJ1327703.pdf

Altarawneh, S., & Hajjo, M. (2018). The acquisition of the English plural morphemes by Arabic- speaking EFL learners. *International Journal of Education & Literacy Studies, 16* (2), 34- 39.

Ardébol, J. (2012) *A morpheme study in a corpus of secondary School EFL* (MA dissertation, Universidad de Granada, Granada, España). Retrieved from https://digibug.ugr.es/handle/10481/22162

Aronoff, M. & Fudeman, K. (2011). *What is morphology?* Blackwell Publishing Ltd.

Booij, G. (2005). *The grammar of words*. Oxford Textbooks in Linguistics.

Britannica, T. Editors of Encyclopaedia (2016, April 14). synchronic linguistics. Encyclopedia Britannica. https://www.britannica.com/science/synchronic-linguistics

Brown, R. (1973). *A First Language.* Harvard University Press.

Callies, M. and Paquot, M. (2015). Learner Corpus Research. An interdisciplinary field on the move. *International Journal of Learner Corpus Research 1*(1), 1-6.

Carlisle, R. (1989). The writing of Anglo and Hispanic elementary school students in bilingual, submersion, and regular programs. *Studies in Second Language Acquisition,* 11, 257-280.

Chuang, F. & Nesi, H. (2008) An analysis of formal errors in a corpus of L2 English produced by Chinese students. *Corpora 1* (2), 251-271. https://www.researchgate.net/publication/250229316_An_Analysis_of_Formal_Errors_in_a_Corpus_of_L2_English_Produced_by_Chinese_Students

Cresswell, J. & Cresswell, J. (2018). *Research Design: Qualitative, quantitative, and Mixed Method Approaches*. Sage.

Demarta, A. (2012). *A study of morpheme order acquisition in an EFL corpus of L1 Spanish- L2 English: Some pedagogical implications.* (MA dissertation, Universidad de Granada, Granada, España). Retrieved from https://digibug.ugr.es/handle/10481/30610

Ellis, N. (2008). The dynamics of Second Language Emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal, 92* (2), pp. 232-249.

Erazo, R. (2022). Analysis of errors in oral production of French L3 learners of level A2 at a high school in Ecuador. *Lingüística en la Red, 19*. Universidad de Alcalá. https://doi.org/10.37536/linred.2022.XIX.1698

Flores, A. & Moore, P. (2023). Diseño y recolección de un corpus oral y longitudinal de aprendices de inglés, Mexican learner corpus (MexLeC). *RLA Revista de Lingüística Teórica y Aplicada, 61*(2), 13-42. http://dx.doi.org/10.29393/rla61-9drap20009

Flores, A. & Moore, P. (2024). Diseño de tareas y materiales para recopilar un corpus de aprendices mexicanos de lengua inglesa. *Estudios de Lingüística Aplicada, 42(*78), 95-123.https://doi.org/10.22201/enallt.01852647p.2024.78.1060

Fontana,U. (2013). *A morpheme order study based on an EFL learner corpus: A focus on the dual mechanism.* (MA dissertation, University of Granada, Granada, España). Retrieved from https://digibug.ugr.es/handle/10481/28284

Granger, Sylviane (2008) Learner corpora en Ludeling, Anke. y Kytö, Merja (Eds.) *Corpus linguistics. An international handbook. Volumen 1* (pp. 259-274). Walter de Gruyter.

Granger, S., Dupont, M., Meunier, F., Naets, H. & Paquot, M. (2020). *The International Corpus of Learner English. Version 3.* Louvain-la-Neuve: Presses universitaires de Louvain. Retrieved from https://dial.uclouvain.be/pr/boreal/object/boreal:229877

Haspelmath, M. (2002). *Understanding Morphology.* Hodder Headline Group.

Hill, J. & Flynn, K. (2006). *Classroom Instruction that works with English Language Learners*. Association for Supervision and Curriculum Development.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in Second Language Acquisition. *Applied Linguistics, 30* (4), pp. 1-22. https://doi.org/10.1093/applin/amp048

Khor, S. (2012). *A corpus based study in morpheme acquisition order of young learners of English: A comparison of Swedish students in grade 6.*

(Dissertation). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-195862

Krashen, S., & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom*. Pergamon Press.

Lieber, R. (2009). *Introducing Morphology.* Cambridge University Press.

Manjunatha, N. (2019). Descriptive Research. *JETIR, 6* (6), pp. 863- 867.

Maslen, R., Theakston, A., Lieven, E., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language, and, Hearing Research. 47*(6), 1319-1333. https://www.researchgate.net/publication/7895534_A_Dense_Corpus_Study_of_Past_Tense_and_Plural_Overregularization_in_English

Mahmood, F., Mahmood, A. & Rasool, A. (2020). A corpus- based comparative study of derivational morphemes across ENL, ESL, EFL learners through ICANALE. *Linguistic Forum - A Journal of Linguistics 2(*4), 1–12. https://www.researchgate.net/publication/351064960_A_Corpus-Based_Comparative_Study_of_Derivational_Morphemes_Across_ENL_ESL_EFL_Learners_Through_ICNALE

McCarthy, A. (2002). An introduction to English morphology. Edinburgh Textbooks on the English Language.

McEnery, T. & Hardie, A. (2012) *Corpus Linguistics: Method, theory and practice.* Cambridge UP.

McEnery, T., Xiao, R. & Tono, Y. (2006) *Corpus-based language studies: An advanced resource book.* Routledge.

Minkkinen, T. (2015). *The influence of morphological congruency in the acquisition of the English plural morpheme -s: A corpus study on Japanese and Spanish learner corpora.* (Dissestion, University of Eastern Finland, Finland, USA).

Muftah, M. (2023). Error Analysis in Second Language Acquisition (SLA): Types and Frequencies of Grammatical Errors of Simple Present and Past Tense in the Elicited Written Production Task of Arab EFL Undergraduate Learners. Colomb. Appl. *Linguistic Journal 25*(1), pp. 42-56.

Nassaji, H. (2015). Qualitative and descriptive research: Data type versus data analysis. *Sage Journals, 19* (2), pp. 129-132.

O'Donnell, M. (2008). *Demonstration of the UAM CorpusTool for text and image annotation*. Universidad Autónoma de Madrid. Retrieved from https://www.researchgate.net/publication/220874508_Demonstration_of_the_UAM_CorpusTool_for_Text_and_Image_Annotation

Palotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics 28* (3), pp. 361-382. https://www.researchgate.net/publication/31354245_An_Operational_Definition_of_the_Emergence_Criterion

Pica, T. (1983). Adult Acquisition of English as a Second Language under Different Conditions of Exposure. *Language Learning, 33*, 465-497.

Polio, C. (1997). Measures of Linguistics accuracy in Second Language writing research. *Language Learning, 47* (1), pp. 101-143.

Sharma, L, Bidari, S., Bidari, D., Neupane, S. & Sapkota, R. (2023). Exploring the mixed methods research design: Types, purposes, strengths, challenges and criticisms. *Global Academic Journal of Linguistics and Literature, 5*, 3-12. https://www.gajrc.com/journal_details/gajll/35/129

Stauble, A. M. (1981). *A Comparative Study of Spanish-English and Japanese-English Continuum: Verb Phrase Morphology.* Unpublished PhD dissertation in applied linguistics. Los Angeles: University of California.

SULEC (2025). *The Santiago University Learner of English Corpus.* Retrieved from https://sulec.cesga.es/

Tognini-Bonelli, E. (2001). *Corpus Linguistics at work*. John Benjamins.

Trinity College London (2025). *Spoken learner corpus*. Retrieved from https://www.trinitycollege.com/about-us/research/Trinity-corpus

Yin,W., & Yang, Y. (2022). Acquisition of English plural making by a Cantonese-English bilingual child: A corpus-based case study. *Journal of Applied Linguistics and Language Research, 9*(2),81-95.

Zhang, S. (1987). Cognitive Complexity and written production in English as a second language. *Language Learning, 37,* 469-481.

Zhang, Y., & Widyastuti, I. (2010). Acquisition of L2 English morphology. *Australian Review of Applied Linguistics*, *33*(3), 29.1-29.17. https://doi.org/10.2104/aral1029