Journal of English Language Teaching
Vol. 68 No. 4 (July-August 2026)
DOI: https://doi.org/10.66121/5382da68

Idiosyncratic Rules among Kurdish Intermediate English Learners' Compositions: A Focus on their Interlanguage

Golaleh Noori[1], Hamid Rahmani Sangani[2] & Somayeh Karimi Haji Pamogh[3]

ABSTRACT

This study investigated recurrent morphosyntactic errors in Sorani Kurdish EFL learners' writing and inferred likely sources for the most frequent patterns at a private language institute in Dehgolan, Kurdistan region in Iran. Using a corpus-based, descriptive design, we analysed 53 learners' writing samples. Errors were identified at the phrase level, coded into categories, and reported as percentage frequencies to determine the most recurring grammatical problem areas and patterns linked to interlanguage development and L1 transfer. The results revealed four main problem areas: articles (44.2%), subject–verb agreement (30.0%), plural marking (20.8%), and pronouns (5.0%). Article errors were the most frequent. Learners often omitted a/an and the, confused the two forms, or used a before plural nouns, all of which can be attributed to Sorani transfer. Subject–verb agreement presented a parallel difficulty: many learners consistently omitted the third-person singular -s (e.g., she give) and selected verb forms based on semantic proximity rather than grammatical subject. Plural errors clustered after quantifiers (e.g., many student), accompanied by count–mass confusion and over-pluralisation of irregular forms. Pronoun errors included resumptive pronouns in relative clauses (e.g., the man that I saw him), object/possessive confusion, and number mismatches in reflexive forms. Collectively, these findings confirm that intermediate learners' grammar remains strongly shaped by the L1 and continues to develop unevenly across morphosyntactic subsystems. This paper addresses a notable gap by providing a morphosyntactic error corpus for Kurdish EFL learners, an under-researched population in Iran, and interprets interlanguage variation within the frame of L1–L2 structural contrast. The results have direct implications for EFL pedagogy, particularly for designing form-focused instructional activities targeting low-salience morphemes.

KEYWORDS: EFL learners; Grammatical morphemes; Interlanguage; L1 transfer; Kurdish Sorani; Error analysis

INTRODUCTION

First-language (L1) influence remains a persistent source of grammatical difficulty in foreign-language acquisition. As Ellis (1997) observed, some of the most enduring barriers to L2 development arise precisely where the L1 and the target language diverge structurally. Even after several months of instruction, learners continue to produce non-target-like forms in areas where the two systems are most misaligned, reflecting what Nunan (2001, p. 89) describes as interference between co-activated grammatical representations. Sorani Kurdish learners of English are a particularly instructive case: the morphosyntactic architecture of Sorani differs from English in ways that are systematic enough to generate predictable error patterns, yet the population has received comparatively little attention in second language acquisition (SLA) research.

Low-salience English morphemes, particularly the plural *-s*, the definite and indefinite articles *a/an/the*, and the third-person singular present *-s*, are well-documented difficulty zones across L1 backgrounds (VanPatten, 1996). Their inconsistent acquisition is partly traceable to the grammatical properties of learners' prior languages. Learners whose L1 does not use free-standing articles tend to rely on definiteness and specificity criteria in unsystematic ways, producing frequent omissions and form substitutions (Ionin, 2006; Ionin, Ko, & Wexler, 2004). These pressures are particularly acute for Sorani speakers. Sorani Kurdish encodes definiteness and number

1 MA in ELT, Velayat state university, Iranshahr, Sistan and Balichestan, Iran ✉ g_nouri65@yahoo.com
2 Assistant professor, Velayat state university, Iranshahr, Sistan and Balichestan, Iran 0000-0002-5098-4402 ✉ H_rahmani_s@yahoo.com
3 MA in ELL, University of Kurdistan, Sanandaj, Kurdistan, Iran, ✉ s.karimi1985hermes@gmail.com

morphologically through nominal suffixes: *-aka* ('the'), *-ek/yek* ('a/an'), *-an* (plural), and *-akan* (definite plural) (Thackston, 2006). Because this system differs structurally from English's article paradigm, it predicts early interlanguage behaviours such as article omission, article substitution, and the production of + plural noun sequences.

The same structural logic applies to verb agreement and pronoun use. Because Sorani marks person and number on the verb itself rather than through an independent inflectional suffix, the English third person *-s* is both typologically unexpected and phonologically weak, making it susceptible to deletion (Thackston, 2006). In the pronominal domain, Sorani employs a single third-person form (*ew*) without grammatical gender, which can delay the stabilisation of the English he/she/it distinction and promote singular *they* or noun repetition as temporary interlanguage strategies (McCarus, 1958; Jarvis & Pavlenko, 2008). Sorani also licenses resumptive pronouns in relative clauses, a construction that translates directly into English as *the man that I saw him*, which accounts for one of the most salient pronoun error types in the present dataset.

Drawing on 53 writing samples from intermediate Kurdish/Sorani EFL learners, this study pursues two objectives: (1) to identify and describe recurrent morphosyntactic error patterns, interpreting them as evidence of cross-linguistic interaction consistent with Jarvis and Pavlenko's (2008, p. 111) position that "the ultimate goal of transfer research is the explanation of how the languages a person knows interact in mind"; and (2) to evaluate whether intermediate proficiency is accompanied by measurable restructuring of the interlanguage, as predicted by developmental stage theories (Gass & Selinker, 1972, 2008). Four research questions guide the investigation: which grammatical features are most frequent in Sorani EFL writing at the intermediate level; how these patterns reflect properties of Sorani transfer; whether higher proficiency produces clear, measurable interlanguage change; and why learners classified as intermediate continue to exhibit elementary level morphosyntactic errors.

The study is organised as follows. Section 2 reviews the relevant literature on error analysis, interlanguage theory, and morphosyntactic difficulty in L2 writing, with particular attention to findings involving Sorani and Kurdish learners. Section 3 describes the research design, participants, data collection procedures, and analytic framework. Section 4 presents the results for each of the four grammatical categories alongside the proposed interlanguage hypotheses. Section 5 discusses findings in relation to the research questions and the broader SLA literature, and Section 6 draws conclusions and outlines directions for future research.

REVIEW OF LITERATURE

Error analysis is a linguistically oriented method for evaluating second-language production by identifying, classifying, and interpreting non-target-like forms (Crystal, 2008, p. 165). Foundational work in the field established that error analysis reveals not merely learner weaknesses but also strengths, windows into the developing grammar that learners are actively constructing (Corder, 1967; Brown, 2000). Brown (2000) drew an influential distinction between *interlingual* errors, which are traceable to L1 influence, and *intralingual* errors, which stem from misanalysis or overgeneralization within the L2 system itself. Both types are informative: the former reveals cross-linguistic structural mismatch, while the latter indicates the learner's hypotheses about the L2.

Global SLA research has extensively examined interlanguage grammar among Chinese, Arabic, Persian, and Turkish learners, producing well-articulated pictures of how the respective L1 architectures shape English development. By contrast, studies focusing on Kurdish EFL learners, especially corpus-based studies, remain sparse. This gap limits how well existing interlanguage models can be generalised to morphologically agglutinative languages such as Sorani Kurdish, whose nominal and verbal morphology differs substantially from the European languages most frequently studied in SLA.

Within a broader theoretical frame, L1 properties systematically shape interlanguage through processes of transfer, regularisation, and processing preference (Odlin, 1989; Jarvis & Pavlenko, 2008; Gass & Selinker, 2008). The contrastive analysis hypothesis (CAH), formulated by Lado (1971) within the behaviourist paradigm, proposed that a principled comparison of the native and target languages could predict areas of difficulty. Although later critiques qualified the strong version of CAH, the core insight that learners struggle most where the two systems diverge has been repeatedly confirmed empirically and continues to inform transfer research. For Kurdish learners, several core properties of English (free-standing articles, number-neutral verb stems, a gendered pronoun paradigm, and gap-strategy

relative clauses) differ sharply from Sorani equivalents, and these are precisely the domains where the most errors are expected and observed.

Interlanguage development is not, however, merely a record of transfer. Gass and Selinker (2008) argue that development is systematic and proceeds through recognisable stages: learners initially omit low-salience morphemes, then overgeneralize rules, then refine their hypotheses toward target norms. Early omission of plural *-s*, third person *-s*, and articles is cross-linguistically robust and reflects attentional limitations on low-communicative-weight morphology rather than simple ignorance of the rule. Overgeneralization, *childrens*, *he goed* ") signals active hypothesis formation. Later stages are considered by morpheme-by-morpheme refinement that is typically non-linear and sensitive to task demands and input frequency (Verspoor et al., 2011).

The small but growing literature specifically on Kurdish EFL learners has so far produced a partial picture. Abdullah (2020) reported that the most frequent problems in Kurdish learners' compositions concerned mechanics, punctuation and capitalisation, while grammatical errors received less systematic treatment. Work more directly relevant to the present study includes Hasan and Ghafur (2011), who documented frequent omissions and misuses of the English article system among Kurdish EFL students, with difficulty at the generic-specific distinction and with zero-article contexts. Omar (2024) extended this line of inquiry to Kurdish university students, attributing many article inaccuracies to L1 influence from Sorani. Sobati et al. (2024) examined Persian-Kurdish bilingual learners' article use within a Mental Spaces framework, linking inaccuracies to both cross-linguistic influence and instructional conditions. More broadly, Mahmood (2016) identified grammar, punctuation, and spelling as primary concerns in writing at two colleges in the Kurdistan Region, and Aljaf, Ahmed, and Salim (2021) found widespread grammatical and lexical errors in a sample of 80 essays from the same region.

None of these studies, however, provides a systematic, multi-domain morphosyntactic profile that simultaneously addresses articles, plural marking, subject–verb agreement, and pronouns within a unified interlanguage framework. The present study addresses this gap directly.

METHODOLOGY

Research Design

This study employs a corpus-based, quantitative, descriptive design. Participants (N = 53) were intermediate Sorani Kurdish learners identified by institute placement tests corresponding approximately to CEFR level B1. Writing samples were drawn from learners' naturally produced texts rather than controlled elicitation tasks, following the naturalistic corpus approach described by McCarthy and Sampson (2005). The research design proceeded through three analytic stages: data collection, error identification and classification, and quantitative frequency analysis. Two qualified coders independently analysed 25% of the corpus, achieving 91% inter-rater reliability before proceeding to full coding. Data triangulation through teacher field notes and coding logs strengthened the findings.

Context and participants

Participants were 53 intermediate Kurdish (Sorani) EFL learners enrolled at a private language institute in Dehgolan, Kurdistan Province, Iran. All participants are native Sorani speakers; many also have contact with Persian as the country's official language, although Persian L2 influence was acknowledged as a potential confound that could not be fully controlled. During a routine class session, participants produced short compositions of approximately 180 words on familiar topics, daily routines, family life, weekend activities, and celebrations, without access to grammar-checking tools and without penalty for language errors. A representative prompt instructed learners to write about their weekend, focusing on ideas rather than accuracy. All submissions were anonymised using ID codes before analysis, and participants were informed that texts would be used solely for linguistic research and not for course assessment.

Units of Analysis and Error Classification

Three complementary units of analysis were employed. First, each sentence was categorised as English-like or non-English-like based on its morphosyntactic profile, following Gass and Selinker (2008). Second, phrase-level segments producing non-target-like morphology were isolated for classification. Orthographic errors were noted but excluded from grammatical metrics unless they bore directly on morphosyntactic analysis. Third, each identified pattern was aligned with a directional hypothesis derived from the SLA literature (Odlin, 1989; Gass & Selinker, 2008; Jarvis & Pavlenko, 2008), and frequency counts and percentages were

computed for each category. Error recognition and categorization followed the taxonomy proposed by Gass and Selinker (2008). Contrasts between Sorani Kurdish and English were informed by reference grammars (Thackston, 2006; McCarus, 1958). Frequency data were analysed and visualised using Excel and SPSS.

LIMITATIONS

The sample comprised learners from a single instructional context, limiting external validity. The dataset consisted solely of brief written compositions; no spoken data or longitudinal observations were gathered. Learner background variables beyond placement test scores were not systematically collected. The study lacked a comparison group at a different proficiency level, and the potential influence of Persian as an additional language could not be fully excluded. The analysis was confined to morphosyntactic indices; discourse-level and broader syntactic phenomena fell outside its scope. The findings are therefore best understood as a practical error profile to inform pedagogy rather than as causal or broadly generalizable claims.

RESULTS

Overall error profile

The corpus yielded 317 morphosyntactic error tokens distributed across four grammatical categories. Article errors were the most frequent (frequency = 140; 44.2%), followed by subject–verb agreement errors (frequency = 95; 30.0%), plural marking errors (frequency = 66; 20.8%), and pronoun errors (frequency = 16; 5.0%). Together, article and SVA errors accounted for 74.1% of all errors. Article errors occurred 1.47 times more often than SVA errors, 2.12 times more often than plural-marking errors, and 8.75 times more often than pronoun errors. A binomial confidence interval analysis confirmed the ranking, establishing article errors as significantly the most frequent category and pronoun errors as the least frequent.

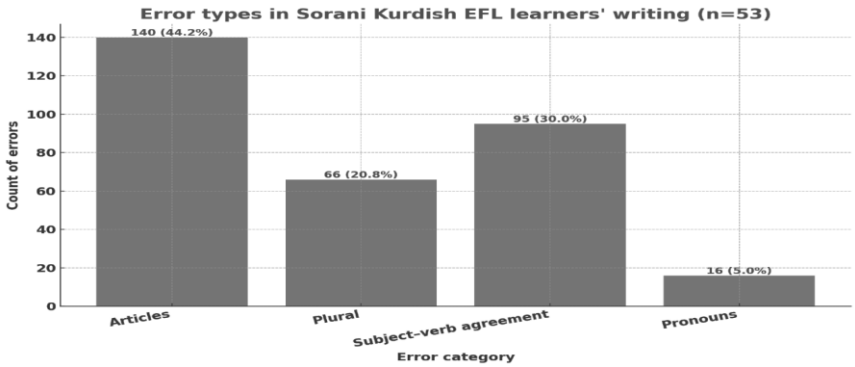

Figure 1. Categories of errors in the writing samples of Sorani Kurdish learners of English as a foreign language

Articles (a/an, the)

Article errors constituted the largest share of the total error count. The error typology (Table 1) identifies 13 error patterns, the overwhelming majority of which involve omissions. English-like article use appeared almost exclusively in superlative constructions (e.g., *the best*, *the most important*), while all other environments – singular nouns, prepositional phrases, generic references, and copular structures – were dominated by non-target-like production. Four interlanguage hypotheses were proposed:

H-A1 (Definiteness Transfer). In contexts requiring a definite or indefinite article before a singular or generic noun, learners frequently omit the article or alternate inconsistently between *a* and *the*. This pattern reflects L1 transfer from Sorani, which encodes definiteness morphologically through the nominal suffixes *-ek* (indefinite) and *-aka/-akan* (definite), rather than through independent pre-nominal articles. The absence of an equivalent free morpheme in Sorani makes the English article category representationally opaque for these learners.

H-A2 (Unstressed Article Deletion). Learners systematically delete articles in rhythmically weak positions, particularly before singular nouns and inside prepositional phrases, producing, for example, *in morning* or *at restaurant* instead of *in the morning* or *at the restaurant*. This pattern, termed Unstressed Article Deletion, suggests that learners prioritise lexically stressed, content-bearing words while treating unstressed function words as dispensable at this stage of development.

H-A3 (Indefinite Article Overgeneralization to Plurals). Learners overgeneralize the singular indefinite article to plural noun phrases, most notably in copular *be* + adjective + noun constructions, yielding forms such as *a good teachers* or *a mothers*. This mismatch between article and noun number

reflects the mapping of English *a* onto the Sorani suffix *-ek*, which marks indefiniteness without requiring any agreement with noun number.

H-A4 (Formulaic Accuracy in Superlatives). Learners produce superlative constructions, *the* + *-est* and *the most* + *adjective*, with near-target-like accuracy. This is interpreted as evidence that highly frequent, structurally prefabricated chunks become entrenched early, providing "islands of accuracy" within an otherwise variable article system.

Table 1. Categorisation of article usage errors

	Non-English-like	Corrected English-like	Error Category
1	*have lots of clever student that they will be a good teachers*	*have lots of clever students that they will be good teachers*	Article misuse with plural (extra *a*) + plural omission
2	*be a good mothers*	*be good mothers*	Article misuse with plural
3	*Do you think sleeping is ordinary manner*	*Do you think sleeping is **an** ordinary manner*	Missing indefinite article (*an*)
4	___	*the best decision I've made for myself*	Correct (English-like)
5	*exercise is the most effective thing in the world that you can do to be healthy*	*exercise is the most effective thing in the world that you can do to be healthy*	Correct (English-like)
6	—	*the most important thing / I like the food from my city the most*	Correct (English-like)
7	*nowruz and Yalda night are the most important traditions that people in IRAN celebrate*	*Nowruz and Yalda night are the most important traditions that people in Iran celebrate*	Capitalisation/convention (near English-like)
8	*to be teacher is important occupation*	*to be **a** teacher is **an** important occupation*	Missing indefinite article (*a, an*)
9	*I have big family*	*I have **a** big family*	Missing indefinite article (*a*)
10	*in Iran new year is called nowroz; Iran is very old country of word*	*In Iran, the new year is called Nowruz; Iran is a very old country in the world*	Missing definite/indefinite article (*the, a*) + word choice
11	*Next day I went to resturant to eat lunch*	*The next day I went to **the** restaurant to eat lunch*	Missing definite article (*the*)
12	*in morning, I went to Dehgolan…*	*in **the** morning, I went to Dehgolan…*	Missing definite article (*the*)
13	*family is the first to help us*	*The family is the first to help us in times of need*	Missing definite article (*the*)

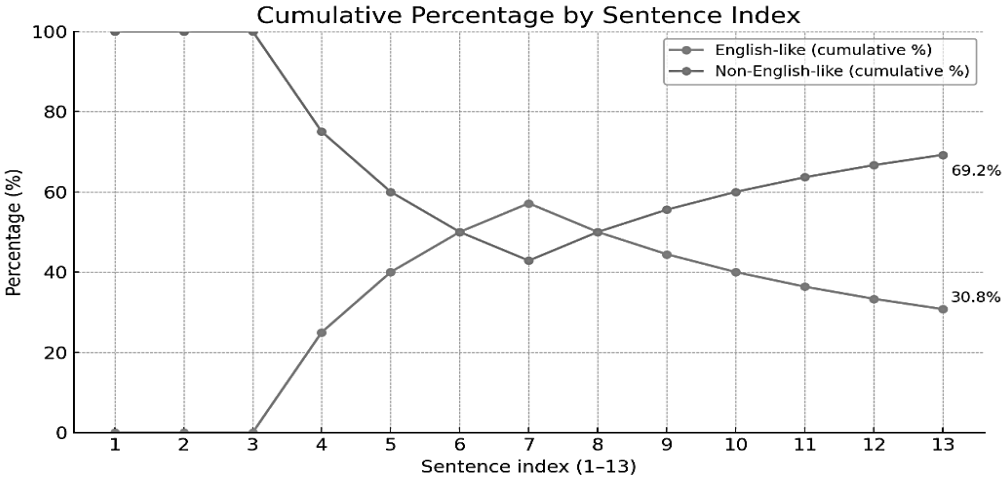

Figure 2. Across items 1–13, article use is predominantly marked by omissions, with English-like patterns largely restricted to superlative forms.

Subject–Verb Agreement

SVA errors formed the second-largest category. Across the 15 items in the error typology (Table 2), SVA production was uniformly non-English-like: the cumulative line for English-like forms remained at zero throughout the entire sequence, indicating that not a single naturally produced SVA context in the classified set was target-like.

Table 2. Categorisation of subject–verb agreement (SVA) usage errors

	Non-English-like	Corrected English-like	Error category
1	"How Rabbit Behave"	"How Rabbits Behave."	Plural morphology (missing **-s**)
2	"Rabbit are small…"	"Rabbits are small…"	Plural morphology (missing **-s**)
3	"I goes to hypermarket"	"I go to the hypermarket."	Subject–verb agreement (SVA); article use (missing **the**)
4	"everyone who have a job"	"Everyone who has a job…"	SVA with **everyone** (singular)
5	"each sports that come…"	"Each sport that comes…"	Determiner–noun number (**each + singular**); SVA
6	"Every country have its own different…"	"Every country has its own different…"	SVA with **every** (singular)
7	"someone need to help"	"Someone needs to help."	SVA with **someone** (singular)
8	"everyone are good people"	"Everyone is a good person."	SVA with **everyone** (singular)
9	"Everything purple attract my attention… everything purple have…"	"Everything purple attracts my attention… Everything purple has …"	SVA with **everything** (singular)
10	"Every sports start in a country… every sport starts …"	"Every sport starts in a country… Every sport starts …"	Determiner–noun number (**every + singular**); SVA
11	"I don't pretending to be ok…"	"I don't pretend to be OK…"	Verb form after **do** (must be base form)
12	"culture mean traditions…"	"Culture means traditions…"	SVA (3rd-person **-s**)
13	"My mom and dad has three children"	"My mom and dad have three children."	SVA with coordinated (plural) subject
14	"Imagining and dreaming has always fascinated me"	"Imagining and dreaming have always fascinated me."	SVA with compound subject (plural)
15	"She give milk but she lays an egg"	"She gives milk, but she lays an egg."	SVA (3rd-person **-s** missing on *give*)

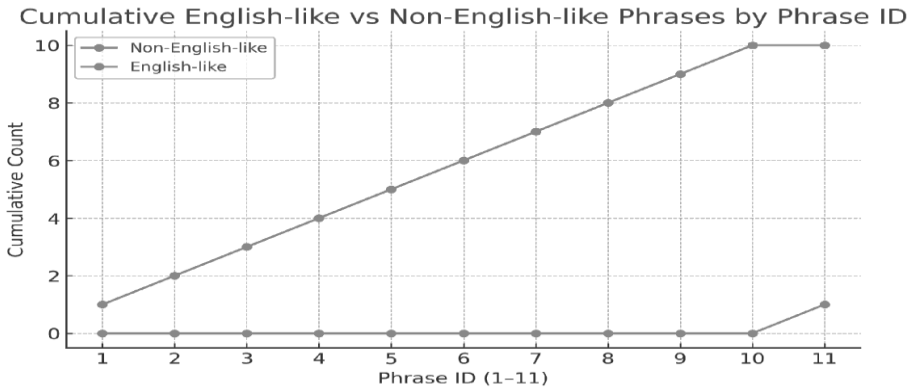

Figure 3. SVA cumulative English-like vs Non-English-like across Ids 1–15. Non-English-like rises monotonically to 15.

Five hypotheses were advanced:

H-S1 (Bare-Verb Default). Learners default to the bare verb stem in the present simple (e.g., *culture mean*, *she give*, *rabbits behave*), supplying the third-person *-s* only under explicit monitoring. This pattern reflects the low perceptual salience and low communicative weight of the *-s* inflection.

H-S2 (Phonological Simplification). The consistent deletion of *-s* in third-person singular present-tense contexts reflects a phonological reduction strategy in which the inflectional suffix is treated as dispensable, consistent with patterns of articulatory simplification observed broadly in early interlanguage.

H-S3 (Semantic Plural Attraction). Despite the grammatical singularity of universal quantifiers – *each*, *every*, *everyone*, *someone* – learners frequently select plural verb agreement, driven by the semantic plurality associated with collective reference. This "plural-attraction" effect is cross-linguistically attested in L1 processing and is amplified in L2 contexts where agreement rules are not yet automatized.

H-S4 (Last-Conjunct Agreement). In coordinated subject constructions, learners match the verb to the nearest (final) noun phrase rather than to the entire coordinated subject. This last-conjunct agreement pattern is a well-documented L1 and L2 processing strategy reflecting syntactic proximity effects.

H-S5 (Double Tense/Negation Marking). Learners overextend do-support, forming past-tense negatives in which tense is marked simultaneously on the auxiliary and the main verb (e.g., *didn't* + *verb*+-*ed*). This pattern is interpretable as a consequence of Sorani's lack of an equivalent do-support construction: learners map past tense morphology twice, once on the auxiliary and once on the main verb, because the L1 provides no template for separating auxiliary and lexical verb in negation.

Plural Marking

Plural errors represented the third-largest category. The 11-item error typology (Table 3) shows that non-English-like forms dominated throughout, with English-like marking appearing only at item 11.

Table 3: Categorisation of plural usage errors

	Non-English-like	Corrected English-like	Error Category
1	Lots of student	Lots of students	Plural omission after quantifier
2	Some country have a lots of big shopping mall	Some countries have a lot of big shopping malls	Plural omission (country, mall)
3	Many country	Many countries	Plural omission
4	Many kind of food	Many kinds of food / Many foods	Plural omission in measure noun
5	Different type of food	Different types of food	Plural omission in measure noun
6	Various sauce like tomato sauce	Various sauces like tomato sauce	Plural omission
7	In different place	In different places	Plural omission
8	Healthy fruit like orange, apple, banana and etc	Healthy fruits like oranges, apples, bananas, etc.	Count vs. mass noun confusion
9	Living room furnitures	Living room furniture	Overpluralisation of uncountable noun
10	Peoples around the world	People around the world	Overpluralisation of collective noun
11	Iran has resources and mines	Iran has resources and mines	Correct (English-like)

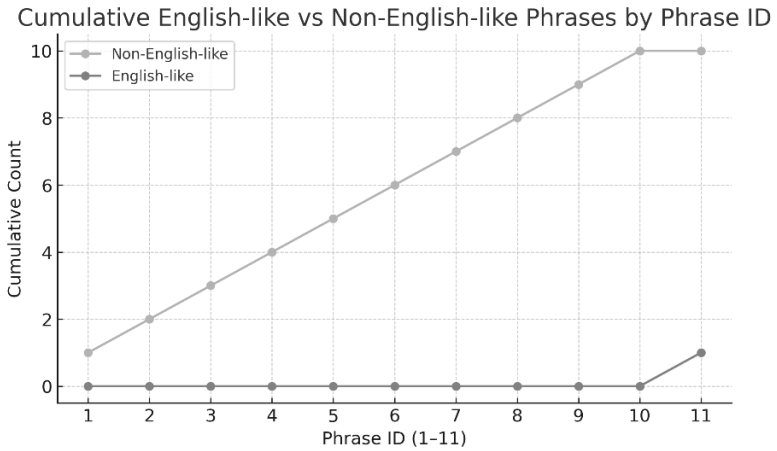

Figure 4: Plurals – cumulative English-like vs Non-English-like across Ids 1–11. Only the final item is English-like; Non-English-like rises to 10.

Five hypotheses were proposed:

H-P1 (Quantifier-Induced Omission). Learners systematically omit the plural suffix *-(e)s* on nouns following quantity expressions, *several*, *many*, *a lot of*, *various*, *some*, *plenty of*, producing forms such as *many student* or *a lot of book*. The quantifier is treated as a sufficient marker of plurality, rendering the nominal suffix perceptually redundant.

H-P2 (Redundancy Perception). This hypothesis extends H-P1: when a quantifier or numeral already encodes plurality, learners generalise the perceived redundancy across all post-quantifier noun contexts, consistently dropping *-(e)s* from count nouns. The error type is therefore not random but reflects a coherent, if non-target-like, interlanguage rule.

H-P3 (Count–Mass Confusion). When learners have not yet acquired the English count–mass distinction for a given noun, they default to treating all nouns as countable, producing forms such as *furnitures* or *informations*. This reflects overapplication of the default count-noun template.

H-P4 (Irregular Plural Overgeneralization). Learners overapply the regular *-(e)s* rule to nouns that require irregular or invariant plural forms (e.g., *peoples*, *informations*), indicating that the default plural rule has been internalised but not yet constrained by lexical exceptions.

H-P5 (Formulaic Accuracy). When a phrase containing a plural noun has been encountered repeatedly as a fixed, positively valenced expression, learners produce correct plural forms (e.g., *Iran has resources and mines*). As with articles, formulaic entrenchment provides a locus of morphological stability within an otherwise variable system.

Pronouns

Pronoun errors, though the least frequent category, displayed clear, systematic patterns. The 13-item typology (Table 4) shows a mid-series recovery of English-like forms, followed by renewed non-English-like production, suggesting that certain pronoun constructions are more tractable than others.

Table 4. Categorisation of pronoun usage errors

	Non-English-like	Corrected English-like	Error category
1	"...people **that you put them** in my life..."	"...people **(that/whom) you put** in my life..."	Resumptive object pronoun in relative clause (extra **them**)
2	"Some people... choose the careers for **her/himself**."	"Some people... choose careers for **themselves**."	Reflexive must agree with plural antecedent
3	"...a lot of money is important for **her/himself**."	"...a lot of money is important for **them / themselves**."	Reflexive with plural referent
4	"I have a friend **her** parents makes her choose her job."	"I have a friend **whose** parents make her choose her job."	Wrong pronoun in relative clause (use **whose**)
5	"...women or man just hurt **their self**."	"...women or men just hurt **themselves**."	Incorrect reflexive form/number
6	"They love me and **I love** their too."	"They love me and **I love them** too."	Incorrect reflexive form/number
7	"...pets depend **themselves owners**."	"...pets **depend on their** owners."	Missing preposition + wrong reflexive problem (depend on + their)
8	"**This** people think..."	"**These** people think..."	Demonstrative number agreement (this → these)
9	people that you put **them** in my life...	...people (that/whom) you put Ø in my life...	Relative clause — **resumptive object pronoun** (extra **them**)
10	friends that I met **them** at school...	...friends (that/whom) I met Ø at school...	Relative clause — **resumptive object pronoun** (extra **them**)
11	the books that I read **them** last summer...	...the books (that/which) I read Ø last summer...	Relative clause — **resumptive object pronoun** (extra **them**)

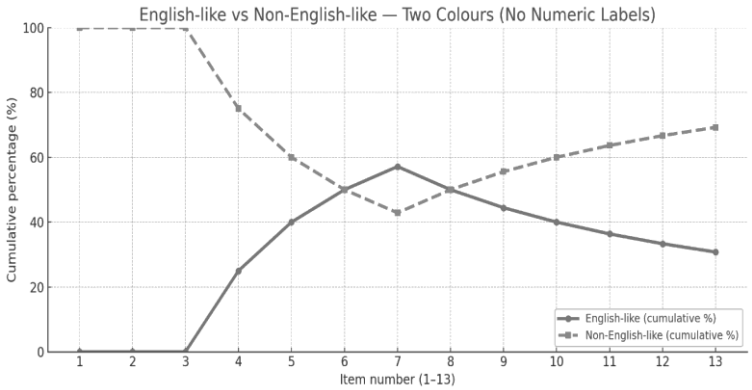

Figure 4. Cumulative share of phrase quality: the blue line (English-like) rises mid-series, while orange (Non-English-like) dominates overall across 13 items.

Three hypotheses were advanced:

H-Pr1 (Resumptive Pronoun Insertion). Learners insert object pronouns (*him*, *her*, *them*) in the position of the relative clause gap (e.g., *people that you put them in my life*; *the man that I saw him*). This resumptive strategy is grammatical in Sorani, which licenses a pronoun copy co-referential with the relative clause head. The pattern represents direct L1 transfer of a typologically different relative clause strategy.

H-Pr2 (Object/Possessive Confusion). Learners substitute object pronouns for possessives and vice versa (e.g., *women or man just hurt their self*), suggesting incomplete acquisition of the morphological boundary between case-marking and possessive functions in the English pronoun paradigm.

H-Pr3 (Reflexive Number Mismatch). When the antecedent noun is plural (e.g., *some people*), learners select a singular reflexive, *himself* or *herself*, instead of *themselves* (e.g., *some people choose the careers for her/himself*). This indicates that the requirement for reflexive pronouns to agree in number with their antecedents has not yet been fully acquired, even when the antecedent number is unambiguous.

DISCUSSION

The four grammatical domains identified in this study, articles, subject–verb agreement, plural marking, and pronouns, together constitute a coherent picture of interlanguage behaviour at the intermediate level. The patterns are neither random nor idiosyncratic in the pejorative sense; they are systematic, rule-governed, and largely predictable from a structural comparison of Sorani Kurdish and English. This consistency strongly supports the L1 transfer hypothesis across all four domains.

Article errors were the most frequent category, and their distribution aligns closely with what Ionin (2006) and Ionin et al. (2004) proposed regarding the role of definiteness and specificity in article acquisition. Learners who have acquired Sorani's suffixal definiteness system face a fundamental representational challenge when encountering English's system of free-standing articles, because the mapping between the two systems is neither direct nor transparent. The result is the omission-dominated pattern observed here, punctuated by the overgeneralization of *a* to plural NPs and by near-target-like accuracy in high-frequency superlative chunks. This "islands of accuracy" phenomenon, in which formulaic expressions achieve target-like form before the underlying rule is mastered, is consistent with models of formulaic language in interlanguage development (Wray, 2002) and with Tarone's (1983) observation that stylistic variation in interlanguage depends partly on the degree to which forms have been automatised.

Subject–verb agreement errors showed the most uniform non-English-like pattern, with zero target-like instances across the classified sequence. This finding is consistent with the established observation that the English third-person singular *-s* is among the last morphemes to be reliably acquired, even by learners at relatively advanced proficiency levels (VanPatten, 1996). From a Sorani-transfer perspective, the absence of this inflection is especially understandable: Sorani encodes person and number agreement directly on the verb stem through a system that bears no functional analogue to English's isolated *-s*. The additional error types, plural-attraction with universal quantifiers, last-conjunct agreement in coordinated subjects, and double tense/negation marking, all reflect processing strategies that prioritise semantic or proximity-based cues over strict syntactic agreement, consistent with the Competition Model's prediction that learners weight cues according to their availability and reliability in the L1 (MacWhinney, 2001).

Plural marking errors clustered most heavily around the post-quantifier context, which is interpretable as a redundancy-driven omission strategy. When learners perceive that the quantifier has already communicated plurality, they assign the nominal -s suffix low informational value and delete it. This is precisely the kind of attentional filtering that VanPatten's (1996) input-processing model predicts: learners prioritise content words and communicatively weighted morphemes, leaving redundant or low-salience inflections unprocessed. The count–mass confusions and irregular plural overgeneralizations in this dataset represent a parallel intralingual phenomenon: the overapplication of default rules before lexical exceptions have been mapped, which is cross-linguistically robust in early interlanguage.

Pronoun errors, while least frequent, were qualitatively revealing. The resumptive pronoun insertion pattern is the clearest single example in the dataset of direct, construction-level L1 transfer: Sorani's licensed resumptive strategy maps directly onto English relative clauses, producing structures

that are entirely grammatical in the L1 but ungrammatical in the L2. The object/possessive confusion and reflexive number mismatch are more likely intralingual in origin, reflecting incomplete acquisition of the English pronoun paradigm's internal distinctions.

Addressing the Research Questions

Regarding RQ1, the four error domains identified, articles (44.2%), SVA (30.0%), plural marking (20.8%), and pronouns (5.0%), confirm that low-salience bound morphemes and free function words are the primary sites of difficulty at the intermediate level. The most frequent individual error types were article omission, bare-verb default in third-person contexts, plural omission after quantifiers, and resumptive pronoun insertion.

Regarding RQ2, the patterns are coherently traceable to Sorani transfer. The suffixal definiteness system predicts article omission and *a* + plural sequences. The absence of a third-person *-s* equivalent in Sorani predicts consistent bare-verb default. The Sorani single third-person pronoun and resumptive relative clause strategy predict both the gender confusion and the pronoun insertion patterns. These predictions are not post-hoc: they derive directly from Thackston (2006) and McCarus (1958) and are confirmed by the corpus data.

Regarding RQ3, the present study design, a single cross-sectional sample without a proficiency-level comparison group, does not permit a direct test of whether higher proficiency produces measurable interlanguage restructuring. The data do show that progress, where it exists, is narrow and domain-specific: learners demonstrate target-like accuracy primarily in formulaic, high-frequency expressions while struggling persistently in productive grammatical contexts. This is consistent with the non-linear, usage-driven developmental trajectory proposed by dynamic systems approaches (Verspoor et al., 2011).

Regarding RQ4, the persistence of elementary-level errors in intermediate learners is explained by several converging factors: the representational distance between Sorani and English morphosyntax, the low salience of the relevant morphemes, the inherent unevenness of interlanguage development (Larsen-Freeman & Cameron, 2008), and the instructional emphasis on communicative fluency at the expense of form-focused feedback. The proficiency label "intermediate" captures overall communicative ability but obscures the morphosyntactic deficit that persists in low-salience domains.

CONCLUSION

This study analysed 317 morphosyntactic errors in the writing of 53 intermediate Sorani Kurdish EFL learners, identifying articles, subject–verb agreement, plural marking, and pronouns as the four primary difficulty domains. The error patterns are systematic and reflect a coherent interlanguage shaped by L1 transfer, low-salience morpheme omission, redundancy-driven deletion, and selective formulaic accuracy. Even at the intermediate level, learners produce morphosyntactic errors characteristic of earlier developmental stages, confirming that proficiency level and morphosyntactic accuracy develop on different trajectories.

The study's theoretical contribution lies in extending interlanguage frameworks, particularly Gass and Selinker's (1972, 2008) developmental stage model and Jarvis and Pavlenko's (2008) transfer framework, to a typologically agglutinative language largely absent from mainstream SLA corpora. The practical contribution is a structured error profile that EFL teachers in Kurdish-speaking contexts can use to prioritise form-focused instruction: specifically, explicit attention to article use in non-superlative contexts, plural -s after quantifiers, third-person agreement in present simple, and relative clause pronoun strategy.

Future research should address the current study's limitations by employing longitudinal designs that track the same learners across proficiency levels, incorporating spoken production alongside writing, and including a comparison group at a higher CEFR level. Controlled elicitation tasks targeting specific constructions would allow finer-grained testing of the interlanguage hypotheses proposed here. Expanding the corpus to include a wider range of task types and sociolinguistic contexts would also strengthen the generalisability of the findings. Taken together, the present results offer both empirical grounding for curriculum design in Kurdistan-region EFL programs and a replicable methodological template for corpus-based error analysis in under-described L1 populations.

REFERENCES

Abdullah, R. (2020). Error analysis in the compositions of Kurdish EFL learners. *Journal of University of Human Development*, *6*(3).

Aljaf, N., Ahmed, S., & Salim, A. (2021). Errors in EFL writing: A study from the Kurdistan Region. *Kurdish Studies Journal*.

Baron, N. (2008). *Always on: Language in an online and mobile world*. Oxford University Press.
Brown, H. D. (2000). *Principles of language learning and teaching* (4th ed.). Longman.
Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, *5*(4), 161–170.
Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.
Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Blackwell.
Ellis, R. (1997). *Second language acquisition*. Oxford University Press.
Ferris, D. (2002). *Treatment of error in second language student writing*. University of Michigan Press.
Gass, S., & Selinker, L. (1972). Language transfer in language learning. *Language Learning*, *22*(2).
Gass, S., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.). Routledge.
Hasan, A., & Ghafur, A. (2011). Kurdish EFL students' article use errors. *Journal of Language Teaching and Research*, *2*(1).
Ionin, T. (2006). This is definitely specific: Specificity and definiteness in article systems. *Natural Language Semantics*, *14*(2), 175–234.
Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 acquisition. *Language Acquisition*, *12*(1), 3–69.
Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.
Lado, R. (1971). *Linguistics across cultures*. University of Michigan Press.
Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford University Press.
Lee, I. (2004). Error correction in L2 secondary writing classrooms. *Journal of Second Language Writing*, *13*(4), 285–312.
MacWhinney, B. (2001). The competition model. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge University Press.
Mahmood, R. (2016). Common writing problems among EFL learners in Kurdistan. *International Journal of English Language Teaching*, *4*(3).
McCarthy, M., & Sampson, N. (2005). *Corpus linguistics: Readings in a widening discipline*. Continuum.
McCarus, E. (1958). *A Kurdish grammar*. American Council of Learned Societies.
Nunan, D. (2001). *Second language teaching and learning*. Heinle & Heinle.
Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.
Omar, A. (2024). Article use errors among Kurdish university students. *Journal of Kurdish Studies*, *12*(1).
Omar, A., & Barzani, S. (n.d.). Writing mechanics errors in Kurdish EFL learners. Cihan University–Duhok.
Sobati, A., et al. (2024). Error analysis of Persian-Kurdish bilingual learners' article use. *TESOL Journal*.
Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, *4*(2), 142–163.
Thackston, W. M. (2006). *Sorani Kurdish: A reference grammar with selected readings*. Harvard University.
VanPatten, B. (1996). *Input processing and grammar instruction*. Ablex.
Verspoor, M., de Bot, K., & Lowie, W. (Eds.). (2011). *A dynamic approach to second language development*. John Benjamins.
Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.