



19th Annual Symposium | September 6-8, 2023

The Effect of Organic Grammar on the Selection of Overgeneralized Forms in L2 English Acquisition and the Role of Literacy

Abdulelah Alsulami
Newcastle University, UK

Abstract

The aim of this study is to further explore learners' overgeneralization of functional morphemes (the use of non-target function words or multi-word strings) during the acquisition of second language (L2) English morphosyntax and explain how the stages of Organic Grammar (or OG stages) affect the choice of overgeneralized forms, as my production data show that my participants use different classes of overgeneralized forms. The results show that the use of certain function words (as placeholders) seems to be peculiar to lower-literate learners or VP-stage learners, while higher OG-stage learners or those with higher literacy use different types of placeholders (e.g., the copula 'be' or personal pronouns).

Keywords: Morphosyntax, Literacy, Overgeneralization

Introduction

The path of language development is argued to be uniform regardless of (1) the learner's native/first language (L1), (2) the learner's age at initial exposure to the target language, (3) the type of exposure, and (4) the learner's educational background (Hawkins, 2001). This conclusion is partly drawn from the morpheme order studies of second language acquisition and the extensive second-language acquisition studies of naturalistic adult immigrants conducted in the 1990s and 2000s. However, the learner's educational background and L1 literacy have received less attention in the field of SLA, whether in generative or usage-based approaches. Using the Noticing Hypothesis, Tarone et al. (2009) examined production data from non-/low-literate L2 learners and argued that literacy has an undeniable effect on the acquisition of morphosyntax. This line of research has important implications for the generative approach, where non-linguistic factors, including literacy, play no role in developing linguistic competence. Vainikka and Young-Scholten (2007) took steps to investigate the role of literacy under their Organic Grammar, which adopts a generative approach and claims that literacy has an indirect role in the development of morphosyntax. However, this role of literacy does not affect the route of acquisition but rather the rate.

More recent studies on low-/non-educated immigrant adults have explored overgeneralization within different theoretical approaches. Applying the Minimalist Program, Julien et al. (2016) tested a claim from previous studies (e.g., Van de Craats & Van Hout, 2010) that L2 non-literate learners use dummy auxiliaries (i.e., overgeneralized forms) during their acquisition of Dutch. Julien et al. argued that learners' misuse and overgeneralization of certain morphemes reflect their developmental stage. The learners used dummy auxiliaries to mark subject-verb agreement and tense before correctly attaching Dutch suffixes to the main verbs. Similarly, Vainikka et al. (2017) examined the acquisition of L2 English by adult learners with no or low levels of L1 literacy. Like Julien et al., Vainikka et al. argued that learners' oral production included overgeneralizations only at specific developmental stages of acquisition beyond the Verb Phrase (VP) stage. The researchers also reported that learners not only used single words as overgeneralized forms but also employed multiword utterances (e.g., 'in the'). In a follow-up study, Vainikka and Young-Scholten (2019) argued that overgeneralization occurs as a result of the learner's search for the relevant syntactic head in the input. Failure to identify the relevant head results in overgeneralization. Using the Basic Variety, a functionalist approach, Mocciaro (2019b) examined the acquisition of L2 Italian morphosyntax by low-educated adult learners. In line with existing L2 research, Mocciaro claimed that learners use overgeneralized forms as soon as they move beyond the Basic Variety. This study seeks to explain how the OG stages and literacy levels affect the choice of overgeneralized forms, as my production data show that participants use different classes of overgeneralized forms. To explore the potential role of literacy in the development of morphosyntax and the use of overgeneralization, the study will first assign stages to the learners, then examine the relationship between OG stages and literacy, and their combined influences on the selection of these forms.

Organic Grammar

Organic Grammar (OG) is a proposal for second-language acquisition that posits that learning a language's grammatical properties involves acquiring its functional projections, with the mental representation of its syntactic structure developing incrementally. As shown in Table

1, OG suggests that language acquisition begins with the bare Verb Phrase (VP), the core or bottom part of a syntactic tree. The expected verbs at this stage are thematic verbs, also known as lexical verbs, which are the main verbs that convey the semantic content of a sentence, representing actions (e.g., 'go') or states (e.g., 'know'), in contrast to auxiliary verbs, which indicate grammatical functions such as tense or agreement. The next stage is the Negation Phrase (NegP), which includes thematic verbs as well as the emergence of the copula 'is'. Following this is the Inflectional Phrase (IP), which includes the Tense Phrase (TP) followed by the Agreement Phrase (AgrP), and finally the Complementizer Phrase (CP) stage. The IP carries grammatical features such as tense and agreement: tense is marked by the bound morpheme -ed on main verbs, indicating the acquisition of the TP, while agreement is marked by the bound morpheme -s on main verbs, or by the copula or auxiliary 'be', indicating the acquisition of the AgrP. This progression occurs as a result of the interaction between Universal Grammar (UG) tools and input. The acquisition of these phrases is implicational; for example, the VP stage precedes the NegP stage, which in turn precedes the TP stage, and so on.

Table 1. Organic Grammar Stages for L2 English/OG Stage Criteria

Stage	Word order	Verb types	Agreement/tense	Pronouns	Syntax
VP	L1 order, then L2 order	Thematic (main) verbs	None	Subject, object pronouns absent	None
NegP	Resembles the L2 apart from complex syntax	Thematic verbs; copula 'is'	None	Pronouns forms begin to emerge	Negation; single clauses; formulaic or intonation-based Qs
TP	Resembles the L2 apart from complex syntax	Thematic verbs, modals; copula forms beyond 'is'	No agreement; some tense, some aspect, but not productive	More pronoun forms, but they can still be missing	Conjoined clauses. Formulaic wh-Qs; yes/no Qs w/o inversion
AgrP	Resembles the L2 apart from complex syntax	Thematic verbs, modals, copula forms beyond 'is'; auxiliaries in all forms and tenses	Productive tense, aspect; some agreement, esp. forms of 'be'	Pronouns obligatory, 'there' and existential 'it'	Simple subordination; wh-Qs but all Qs may lack inversion
CP	Always resembles the L2 thematic	Complex tense, aspect forms; passives; range of thematic verb, modal, auxiliary forms	Forms usually correct, apart from newly attempted ones	Use of 'there' and 'it' beyond stock phrases	Complex subordination. All Qs with inversion

Methodology

As mentioned in the introduction, this study aims to explore how OG stages influence the choice of overgeneralized forms and examine the combined role of literacy in these choices. Therefore, this section investigates the acquisition of morphosyntax to assign OG stages using the theory of Organic Grammar and assesses reading levels based on Ehri's (1994) criteria as a measure of literacy to determine the participants' literacy levels. During the tasks, the data were recorded and later transcribed orthographically, with each transcription reviewed twice to ensure accuracy, particularly in capturing the learners' use of non-target elements. The transcribed data were then entered into Microsoft Excel and subsequently imported into SPSS for statistical analysis.

Participants

As shown in Table 2, data were collected from 60 learners (33 males, 27 females) who were low- to moderately educated, meaning they did not complete their compulsory education. All participants were Arabic speakers, with the majority coming from Syria. The learners' ages ranged from 19 to 60 years, and their length of residence in the UK varied from 8 months to 9 years. Their ESOL instruction ranged from none to 5 years, and their native schooling ranged from 2 to 11 years.

Table 2. Descriptive Statistics

Variable	N	%	Mean	Min.	Max.
Age			35	19	60
Sex					
...M	33	55%			
...F	27	45%			
Length of UK residence				8 months	9 years
ESOL instruction				0	5 years
Native schooling				2 years	11 years

Reading Assessment as A measure of Literacy

To determine whether the participants could read and write in Arabic, I asked them to read the first few lines of a story used for the word awareness tasks. I also asked them to write my name and five Arabic words to confirm their schooling. For the reading test, I followed Young-Scholten and Strom (2005), who used various tests adapted from other assessments to evaluate a variety of basic sub-skills in reading, as shown in Table 3. This battery of tests included presenting the learners with the alphabet in an unordered sequence, in different fonts, and with survival/environmental signs. To assess the learners' decoding skills, they were asked to read a list of ten high- and low-frequency monosyllabic and multisyllabic words they might encounter in their everyday lives. These words ranged from high-frequency words, such as 'table' and 'community' to low-frequency words. The selection of low-frequency terms was based on the assumption that these words would not be included in the sight word repertoires of low-literacy ESL students and would thus reveal whether learners had decoding skills. Given the emphasis on phonological awareness and basic reading skills, reading comprehension was not tested.

Table 3. Reading Tests, partly adopted from Young-Scholten and Strom (2005: 54)

Tasks in native language	Tasks in English
<ul style="list-style-type: none"> • Read part of a story • Write five Arabic words 	<ul style="list-style-type: none"> • Read 10 varied single letter identification • Read 4 survival signs • Read a paragraph • Read 10 isolated words

Tasks for Assigning OG Stages

This section presents the tasks used for assigning OG stages to participants. These tasks include investigating L2 English word order, which constitutes the VP stage, and English negation, which constitutes the NegP stage. Additionally, the tasks explore the acquisition of functional morphology, such as the irregular or regular past tense *-ed*, an important morphosyntactic element for assigning a TP stage to participants, and the acquisition of agreement features for the copula and auxiliary ‘be’ and the third-person singular *-s*, which are crucial for assigning an AgrP stage to learners.

Since the focus was on eliciting spontaneous speech, no practice items were provided before data collection began. Learners were familiarized with the nature of each task by indicating the target tense, as will be discussed in each of the following tasks. Due to the expectation of low morphosyntactic performance, no discontinuation rules for errors were applied. Instead, a three-time cut-off criterion was used to determine if the learner had acquired the target morphosyntactic structure. This criterion fits my data sample, as if I had used more target-like productions, very few learners would have acquired TP. This method also distinguishes between learners who prefer to use bound morphemes (the regular past tense *-ed* or the third-person singular *-s*) and those who do not. This criterion could reveal the early stages of functional projections (TP and AgrP).

VP Task

In this task, the learners were introduced to a series of eight picture sheets. For the first picture sheet, when I showed the picture, I said the beginning of each sentence (e.g., ‘the girl...’), paused, then indicated that they should continue and finish the sentence. Because the focus of this task was on word order patterns, the learner was not expected to produce a specific tense (e.g., simple present, present progressive) or specific verbs, nouns, or object names for items in the picture.

NegP/Negation Task

To elicit data that would help investigate negation, the learners were presented with four groups of pictures, each consisting of two similar pictures. The first picture contains an action, while the second one does not. For example, the first picture shows a girl washing a car, and the second shows a girl just holding a sponge. I asked the learners to explain the difference between the two pictures. Understanding the intended purpose of this task might be challenging for some learners. Therefore, I prompted them to say ‘no’. Participants were expected to produce sentences such as ‘The girl doesn’t wash the car’. In this task, the learners were not required to focus on a specific tense, but they had to avoid structures such as ‘no + V’ or ‘no + V + ing’. If they used a structure like ‘no + V’, I asked them to say a full sentence.

AgrP Task: Third Person Agreement on Auxiliaries (Progressive)

The learner saw a set of eight pictures with progressive actions, and I asked the learner questions. For example, I asked, ‘What is this girl doing?’ I also implied that the action was happening now using words and phrases like ‘now’ or ‘at the moment’. The participant’s task was to supply a target syntactic form such as ‘be + V-ing’, as exemplified in (1).

(1) The girl is reading a book.

AgrP Task: Third Person Agreement on Main Verbs

I presented the learners with four groups of pictures that depict habitual actions, each featuring the same character to indicate that this person has habitual actions. To coax the learners to focus on the intended syntactic form, I first told the learner that the person in the picture ‘does the same thing every day’. The learner was expected to supply syntactic constructions such as ‘S + V-s + O’, as shown in (2). The prompts are in square brackets [...] to indicate habitual actions.

(2a) [Every morning], this man drinks coffee.

(2b) [Then], he smokes a cigarette.

TP (-ed)/Tense Task

To investigate TP (-ed), the participants were presented with a collection of sequential pictures showing an incident where a man and a woman saw a boat hit a bridge. The participants were then asked to re-tell the story. To help the participant focus on the past tense, I used temporal adverbials (e.g., ‘last year’) to indicate that the action happened in the past. The learners were expected to use a construction that attached the verbal suffix -ed to the main verb (V + ed), as in (3a), or an irregular past simple verb, as in (3b).

(3a) They watched a boat on the river.

(3b) Then the boat sank into the river.

AgrP Task: Agreement on the Forms of the Verb ‘be’

The aim of this task was to elicit data related to the acquisition of the copula (‘is’ and ‘are’). The learners were shown eight cards (four sets of two cards each). The first card depicted a person or two people with a profession (e.g., a picture of a doctor). To help the learners understand the purpose of this task, I asked them in Arabic to explain what was on the card in English (e.g., ‘he is a doctor’ or ‘they are doctors’). The learners were expected to produce a sentence using the target copula be (‘is’ and ‘are’) according to the card presented.

Results

This section will first determine a reading level for participants and assign OG stages, and then explore the potential relationship between literacy and OG stages and their combined role in the selection of overgeneralized forms.

Reading Results

Based on learners’ results on the five reading sub-tests for English provided earlier in the previous section, I assigned each learner to one of the levels or implicational stages of reading

displayed in Table 4. All learners performed well on varied single-letter identification; scores below 100% were attributable to ‘p’ and ‘b’ or ‘v’ and ‘f’ phonology-based non-discrimination. The survival/environmental sign task had the second-highest score, followed by paragraph reading and decoding. The distribution of these levels will be presented below when discussing the relationship between literacy levels and their corresponding OG stages.

Table 4. Reading level scoring, based on % correct on English reading tests

Reading levels	Varied single letter identification	Survival signs	Paragraph reading	Decoding of familiar words in isolation
1	75%+	25%+	no ability	0%
2	75%+	75%+	attempt	20%+
3	100%	100%	slow, sometimes accurate	20%+
4	100%	100%	mostly accurate	60%+
5	100%	100%	fluent	100%

OG Stages

Based on the OG criteria for L2 English, learners were placed in different OG stages. For example, learners who were placed in the bare VP stage produce utterances with only an optional specifier (the subject), bare verbs, and their complements (e.g., the direct object). Under Organic Grammar, the copula ‘be’ and the auxiliary ‘be’ are base-generated in INFL and are not expected to be produced in the VP stage. However, Vainikka and Young-Scholten (1996b) reported similar examples from two VP-stage learners who produced many copulas (exclusively *is(t)* ‘is’, apart from one instance of *bin* ‘am’). Similarly, the analysis in this study placed three learners in the VP stage who used the copula ‘be’ five times in their attempts. This is because (1) the use of the copula ‘be’ emerged as a default form (e.g., ‘they is student’), and (2) the use of the copula ‘is’ followed pronominal subjects, which suggests that rote learning is involved. Both reasons indicate that the copula ‘is’ is unanalyzed. This use of the copula ‘be’ does not indicate that they have fully acquired the agreement paradigm for the copula ‘be’ at this stage.

The learners who were placed in the NegP stage not only produced the copula ‘is’ but also produced the copula ‘are’. For example, most of these learners at this stage consistently used the copula ‘is’, and some of them used the copula ‘are’. This indicates that some of these learners appear to have acquired the agreement paradigm for the copula ‘be’ because most of these learners who produced the copula ‘are’ also produced the copula ‘is’. However, because they do not show any evidence of TP (such as using regular or irregular past tense) or fully-fledged agreement (like the third-person singular *-s* on the main verbs), they were placed in the NegP stage.

The learners in the TP stage have presented evidence that demonstrates their acquisition of TP by producing the past tense. Their data also suggest that they have acquired VP and NegP. In contrast to learners in the previous stage, these learners demonstrate greater consistency in using both the copula ‘be’ and the auxiliary ‘be’ but not the third-person singular *-s*. This indicates that the agreement feature for free morphemes has begun to be established. However, it should be noted that there are still some learners in this stage who may occasionally omit the auxiliary ‘be’. This is to be expected, as their acquisition of full-fledged agreement is still in progress.

The learners in the AgrP stage appear to have acquired the full agreement paradigm. For example, they have shown evidence of acquiring the agreement feature for free morphemes (the copula ‘be’ and the auxiliary ‘be’) as well as the agreement feature on the main verb. Unlike the learners in the previous stage, these learners consistently produce the copula ‘be’ and the auxiliary ‘be’ correctly. They have also demonstrated evidence of acquiring VP, NegP, and TP.

Examining the Relationship Between Literacy Levels and OG Stages

The initial investigation of the relationship between literacy and OG stages, as shown in Table 5, indicates a noticeable increase in the mean values. This suggests that learners with higher literacy levels tend to attain higher OG stages. Additionally, the SD indicates that learners with higher literacy levels are more likely to reach these stages consistently, as the SD is significantly lower compared to those with lower literacy levels.

Table 5. Mean and SD for OG Stages across Literacy Levels

Literacy levels	N	Mean	SD
1	14	1.43	0.51
2	13	1.69	0.75
3	21	2.57	0.75
4	12	3.08	0.79

The scatter plot in Figure 1 provides further support for this positive relationship. It reveals a positive monotonic relationship between OG stages and literacy levels. In this figure, OG stages represent the learners' progression in the development of morphosyntax, while literacy levels refer to their literacy proficiency. Each dot in the figure represents combined data for a particular literacy level and OG stage. Thus, an increase in literacy levels is accompanied by a corresponding increase in OG stages.

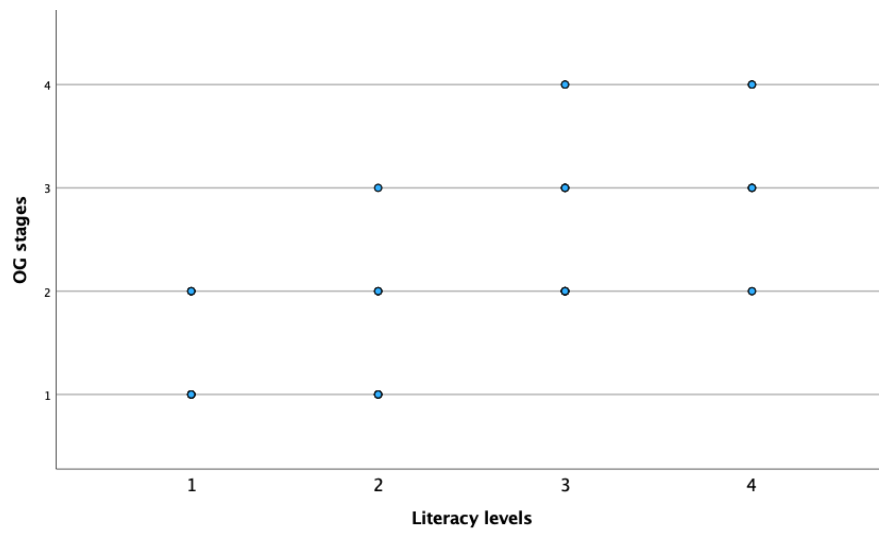


Figure 1. Literacy Levels and OG Stages

To confirm whether this relationship is significant, Spearman's rank correlation was utilized to examine the correlation between OG stages and literacy levels, both of which are ordinal variables (e.g., OG stages being assigned values of 1, 2, 3, 4, and similarly for literacy levels).

Spearman serves as an equivalent non-parametric test when the data fail to meet the assumptions of the parametric Pearson test (e.g., linearity or normality of the data). Spearman does not impose assumptions beyond the requirement that the data be at least ordinal. As Table 6 shows, there is a strong positive correlation between OG stages and literacy levels, as indicated by a significant p-value of 0.01. This means that as literacy levels increase, we can expect an increase in OG stages. The reliability test indicates a high level of consistency, as shown by a Cronbach's Alpha coefficient of 0.798.

Table 6. Correlation between OG Stages and Literacy (Spearman’s rank correlation coefficient)

	Correlation	P-value
Literacy	0.681	0.001

Overgeneralization by Adult L2 Immigrants

This section presents an in-depth analysis of overgeneralizations observed in some learners' production data from different OG stages. The analysis includes all learners who used overgeneralizations, allowing the study to investigate how OG stages affect the choice of overgeneralized forms.

Overgeneralization at the Early Stages of Organic Grammar.

As previously discussed, the analysis placed 13 learners in the VP stage because their VPs contain an optional specifier (the subject), bare verbs, and their complements (e.g., the direct object) with no functional morphology, apart from the copula ‘is’, which is also used as a default form (e.g., ‘they is students’). Their negation construction contains only ‘no+ verb’ or multi-word strings (such as ‘I don’t know’, ‘I don’t like’, or ‘I don’t’) within utterances (see examples below). Four of these learners used overgeneralized forms, as shown in Table 7. To capture the syntactic behavior of the copula ‘be’ as well as NegP development more explicitly, I categorized learners’ production into three categories: ‘is’, ‘are’, and default ‘is’. I also organized their NegP production into four categories based on negation constructions and NegP multi-word chunks (e.g., ‘I don’t know’). This categorization aimed to help identify the exact OG stage where learners used single function words (e.g., ‘for’) or multi-word strings (e.g., ‘you have’) as overgeneralized forms, which were observed in their production. OG stages organize rows in all tables. The table also includes three learners from the NegP stage, but their performance will be discussed further below.

Table 7. Relationship between the acquisition of copula be and the use of overgeneralized forms

Name	OG stage	Copula <i>be</i>			Over-generalized forms	NegP					NegP chunks
		<i>Is</i>	<i>Are</i>	Default <i>Is</i>		No	Not	Is+not/i sn’t	Don’t	Doesn’t	
Omar	VP	0	0	0	0	0	0	0	0	0	8
Mansoor	VP	4	0	4	4	0	0	0	0	0	8
Arshad	VP	4	0	4	11	0	0	0	0	0	8
Adeeb	VP	0	0	0	34	8	0	0	0	0	0
Zaid	NegP	4	0	4	7	0	0	0	8	0	0
Gharam	NegP	0	0	0	24	0	0	0	8	0	0
Sameer	NegP	4	2	3	3	0	0	0	6	0	2

As can be seen in this table and as the examples in the following sections show, none of these VP-stage learners had fully mastered all the patterns of negation structures. Table 5 shows a negative relationship between the overgeneralized use of single function words and multi-word strings and the acquisition of forms of the copula ‘be’. For example, learners who produced overgeneralized forms have not yet acquired the copula ‘be’ or used it as a default form (e.g., ‘they is students’). In the context of the copula ‘be’, two learners (Omar and Adeeb) showed no use of the copula verbs ‘is’ and ‘are’ and did not substitute them with any lexical items (e.g., the preposition ‘for’ or multi-word strings), as shown in (4). Instead, they omitted the copula entirely.

(4) Woman teacher. (Adeeb)

Target utterance: She/The lady is a teacher.

As previously mentioned, the table also includes three NegP-stage learners who used overgeneralized forms in their production data. The analysis shows that these three learners seemed to start developing their NegP projection, which is indicated by the appearance of the sentential negator ‘don’t’. Now, let us move on to discussing the examples of these overgeneralized forms. First, I will discuss the occurrence of single function words, followed by multi-word strings in declarative sentences and multi-word strings in the development of negation (first in sentences with verbs and then in verb-less utterances).

Overgeneralization of Single Function Words. In this section, I present representative examples of single function words used as overgeneralized forms. I will begin with the first learner at the VP stage (Mansoor), who produced the pronoun ‘I’ five times in his utterances between the subject and the lexical verb, as shown in (5). The use of ‘I’ as a default overgeneralized form—unlike the other personal subject pronouns discussed further below, which are used by higher OG stage learners—suggests that those learners have acquired grammatical features such as subject-verb agreement. Despite their varied non-target uses, these pronouns are considered function words because they primarily encode grammatical information, such as agreement, and their use as overgeneralized forms appears to serve morphosyntactic functions.

(5) This baby I drive care. (Mansoor)

Target utterance: The baby drives/is driving a car.

Another example of single function words used as overgeneralized forms is by Adeeb, whose production data includes non-target function words, particularly the preposition ‘for’. As shown in (6), the sentences that Adeeb produced in all tasks, apart from the copula ‘be’, either in declarative or negative contexts, contain single function words.

(6a). The girl for drink Coca Cola. (In the VP task) (Adeeb)

Target utterance: The boy drinks/is drinking cola.

(6b). Children no for clean. (In the NegP task)

Target utterance: The boy doesn’t clean/is not cleaning the room.

(6c). Three boys for clean dishes. (In the present progressive tense -ing task)

Target utterance: Three boys/the boys are cleaning the dishes.

(6d). Every evening boy for clean table. (In the present third-person singular -s task)

Target utterance: Every evening, the boy sets the table.
 (6e). Boat for water. (In the TP task)

The examples of the non-target function word ‘for’ in Adeeb’s speech fragments do not always align with Lakshmanan’s (1993/94) claim that it solely serves as a case assigner. Adeeb uses ‘for’ both when he knows the verb (as a case assigner), as in (6a), and when he does not know it, as in (6b). This indicates that the verb Adeeb uses in some of his sentences already functions as a case assigner.

Gharam, a NegP-stage learner, also used single function words as overgeneralized forms. She overused the pronoun ‘his’ in nearly all tasks, except for the copula ‘be’ and the TP tasks, as shown in (7). One might think that her use of this pronoun is simply a mispronunciation of the copula form ‘is’, but she did not use ‘his’ in the copula ‘be’ tasks. She used ‘his’ three times in combination with the negator ‘don’t’, as in (7d). She consistently used the negator ‘don’t’, again as in (7d), regardless of the required agreement feature, indicating that she overgeneralizes ‘don’t’ to ‘doesn’t’.

- (7a) The girl his eat fish. (In the VP task) (Gharam)
Target utterance: The girl eats/is eating fish.
- (7b) Son his play cat. (In the present progressive tense -ing task)
Target utterance: The girl is carrying a cat.
- (7c) Man Monday his go to shopping. (In the third-person singular -s task)
Target utterance: On Mondays, the man goes to the supermarket.
- (7d) Son his don’t clean house. (In the NegP task task)
Target utterance: The boy does not/is not cleaning the room.

Multi-word Strings. Similar to what has been observed in the literature discussed previously regarding the use of multi-word chunks, three learners at the early stages (VP and NegP) produced them. For example, the VP-stage learner (Arshad) used the phrase ‘in the’ after the main verb, as in (8a), and between the subject and the object, as in (8b). The example in (8c) clearly shows the exact syntactic position of ‘in the’ as it appears before the main verb (‘clean’).

- (8a) The man drown (draw) in the brush. (In the VP task) (Arshad)
Target utterance: The boy paints/is painting a chair.
- (8b) The woman in the brush in dog. (In the VP task)
Target utterance: The girl is washing/washes a dog.
- (8c) My children in the clean sink. (In the present progressive tense -ing task)
Target utterance: The children are cleaning the dishes

Another instance of chunks comes from Zaid, a NegP-stage learner, who produced the phrase ‘you have’ in a couple of his utterances before the verb, as in (9a), and in his verbless utterances, as in (9b). The phrase ‘you have’ also appears in Sameer’s data, as in (9c).

- (9a) Sarah you have eat fish. (In the VP task) (Zaid).
Target utterance: Sarah eats/is eating fish.
- (9b) Layla you have horse.
Target utterance: Layla draws/is drawing a horse.

(9c) The baby you have drive car. (In the VP task) (Sameer)

Target utterance: The baby drives/is driving a car

Multi-word Expressions in the Development of NegP. The analysis shows that three VP-stage learners (Arshad, Mansoor, and Omar) and one NegP-stage learner (Sameer) produced such chunks in the development of their negation constructions. As the data in Table 8 illustrates, these learners used three types of NegP chunks: ‘I don’t know’, ‘I don’t like’, and ‘I don’t’. The table also shows the frequency of these chunks.

Table 8. NegP chunks

Name	OG stage	NegP chunks		
		‘I don’t know’	‘I don’t like’	‘I don’t’
Arshad	VP	7	0	1
Mansoor	VP	0	8	0
Omar	VP	6	2	0
Sameer	NegP	0	0	2

The rest of this section introduces some representative examples of NegP chunk use. I will begin with the use of the unanalyzed multi-word string ‘I don’t know’ produced by two VP-stage learners (Arshad and Omar). These two learners used it between the subject and the verb for negative phrases, as in (10).

(10) Boy I don’t know clean house. (Arshad)

Target utterance: The boy does not clean/is not cleaning the house.

Another example of the production of chunks is the multi-word string ‘I don’t like’, which two VP-stage learners (Mansoor and Omar) used. Again, ‘I don’t like’ was used between the subject and the verb, as in (11).

(11) This girl I don’t like clean car. (Mansoor)

Target utterance: The girl does not wash/is not washing the car.

The last example of chunks shows the use of the multi-word string ‘I don’t’. Sameer used this multi-word string twice before the verb, as in (12a). This multi-word string also appeared only once in Arshad’s utterances, as in (12b).

(12a) Boy I don’t drive car. (Sameer)

Target utterance: The boy doesn’t/is not driving a car.

(12b) Girl I don’t clean and car. (Arshad)

Target utterance: The girl doesn’t wash/isn’t washing the car.

NegP Chunks in Verbless Utterances. Just as in the cases of the production of function words in verb-less utterances, further analysis of the data shows that there are a couple of verb-less utterances produced by learners that contain the unanalyzed multi-word string ‘I don’t know’, as in (13). These strings often precede the object.

(13) Boy I don't know a car. (Arshad)

Target utterance: The boy doesn't wash/isn't washing the car.

Overgeneralization at Higher Stages of Organic Grammar

According to the analysis in this paper, some learners from the NegP (n = 7), TP (n = 4), and AgrP (n = 2) stages used overgeneralized 'be' and/or personal subject pronouns either in the same context or in different contexts. Unlike the NegP-stage learners who used single function words or multi-word strings, these learners are more advanced, as indicated by their consistent use of the copula 'be' and their very limited use of auxiliary 'be'. The learners at the TP stage, in addition to their consistent use of the copula 'be', started to use auxiliary 'be' and, to a very limited extent, used the past tense or the third person singular -s. The learners at the AgrP stage, in addition to their consistent use of the copula 'be', used auxiliary 'be', the past tense, and some third person singular -s. This indicates that OG stages affect the choice of overgeneralized forms.

The rest of this section presents some representative examples of these types of overgeneralizations. I will first present examples of the overgeneralization of 'be', followed by examples of the overgeneralization of personal pronouns. For example, the NegP-stage learner Nihad used 'is' before the main verb, as in (14a). Another example comes from Maher, at the TP stage, who used 'are' before the main verb, as in (14b). These examples suggest an agreement with the subject in number as 'be-forms' change according to the subject.

(14a) The baby is drive a car. (In the VP task) (Nihad)

Target utterance: The baby drives/is driving a car.

(14b) The kids are play game. (In the VP task) (Maher)

Target utterance: The kids are playing video games.

Moving on to the use of personal pronouns or a combination of personal pronouns and 'be-forms' as overgeneralized forms, the analysis shows that five learners used subject personal pronouns in this way; four of them are at the NegP stage, and one is at the AgrP stage. Unlike the pronoun 'I' or 'his' discussed earlier in this section, which was used as a default form, these pronouns clearly indicate a morphosyntactic function. For example, these pronouns change according to the subject, agreeing with the subject in gender, as in (15a) or number, as in (15c).

(15a) Sarah she read the book. (In the present progressive tense -ing task)

Target utterance: Sarah is reading a book.

(15b) The boys they writing letters. (In the present progressive tense -ing task)

Target utterance: The boys are writing letters.

These examples indicate an effect of L1 functional morphology. However, OG dismisses any role of L1 in the acquisition of L2 functional morphology.

For the combination of personal pronouns and 'be-forms' as overgeneralized forms, there are also five learners who used this type of overgeneralization. For example, Afnan, who is at the NegP stage, used the pronoun 'he' and the verb 'is' in her sentences, as in (16). This indicates an effect of L1 functional morphology but also suggests that 'is' and 'he' have different syntactic functions. The 'is' marks agreement, whereas the pronoun 'he' marks gender. In Arabic, lexical verbs are inflected for person, number, and gender. If this is correct, it means that L1 transfer

occurs in the intermediate stages (e.g., TP or AgrP stage where learners acquire functional morphology) (Hawkins, 2001).

(16a) The boy he is clean his bedroom. (In the VP tasks) (Afnan)

Target utterance: The boys/They are cleaning the dishes.

(16b) Emily she is have her breakfast. (In the third-person singular -s tasks)

Target utterance: Emily eats her breakfast.

Discussion

Do the Stages of Organic Grammar Influence the Choice of Overgeneralized Forms? If So, How?

The type of overgeneralized forms appears to reflect the learners' stage of development (i.e., OG stage). The learners in this study used sub-patterns of overgeneralized forms. As noted in the previous section, learners at the VP stage used single function words (e.g., 'for') or multi-word strings as overgeneralized forms. These learners, who used these forms, have not yet fully developed their acquisition of NegP and the copula 'be'; therefore, they were assigned to the bare VP stage. They also used expressions such as 'I don't know', 'I don't like' or 'I don't' in the task aimed at eliciting negation. For the copula 'be', they often used 'is' as a default form (i.e., they overgeneralized 'is' to 'are'). Learners with higher OG stages used the copula 'be' and personal pronouns as overgeneralized forms. I believe this choice of overgeneralized forms is ascribed to the nature of their stage. Specifically, the bare VP stage lacks any IP-related elements (e.g., the copula 'be', auxiliary 'be' or bound functional morphemes such as the past tense marker -ed or the third-person singular -s); therefore, learners at this stage are more likely to use other forms they were previously exposed to in the input. These forms change according to the elements available in the subsequent stage. As noted in the previous section, the learners who were placed at the stage of NegP or TP used be-forms or subject personal pronouns as overgeneralized forms, which are available at these stages.

The choice of overgeneralized forms has also been observed by Mocciaro (2019b), who reported that the overgeneralization of *fare* 'to do' is connected to the transition from the basic to post-basic varieties, whereas the overgeneralization of *essere* 'to be' may only appear at a slightly more advanced stage when learners are in the process of acquiring the copula 'be'. This pattern aligns with the prediction of placeholders in L2 English, where the copula 'be' is a candidate as a placeholder for the head of TP or AgrP. However, my data and this prediction contradict Julien et al.'s (2016) findings regarding the use of dummy *zijn* 'be' and *gaan* 'go'. Julien et al. claimed that learners with lower levels of language acquisition used both dummy auxiliaries, while more advanced learners continued to use *gaan*. They proposed two phases concerning dummy use: an initial phase in which both dummy *zijn* and *gaan* are used, followed by a phase in which dummy *zijn* occurs less frequently, but dummy *gaan* continues to be used. I believe this assumption should be reconsidered, as Julien et al. only investigated whether *zijn* and *gaan* represent syntactic functions. Their assumption is based on language proficiency rather than on a theory of second language development (e.g., the Basic Variety or Organic Grammar), which can determine the specific stage or point in language development when and where overgeneralized forms begin to appear and disappear.

The Role of Literacy in Morphosyntactic Development and the Use of Multi-Word Strings

As shown in the results, there is a strong positive correlation between OG stages and literacy levels. This suggests that learners with lower literacy levels are likely to exhibit lower levels of morphosyntax, corresponding to lower OG stages. This finding points to a significant role of literacy in the development of morphosyntax, potentially having an indirect effect on language acquisition. Literacy affects the development of morphosyntax by strengthening metalinguistic awareness, which supports the recognition and processing of grammatical structures. In instructed classroom settings, learners often exhibit faster language development compared to naturalistic learners. Researchers attribute this difference to the metalinguistic information that learners receive in the classroom. In traditional classroom environments, teachers provide explicit instruction on grammatical properties, followed by exercises to reinforce learning. For example, learners are explicitly taught that the morpheme *-s* is attached to verbs to mark agreement with singular subjects in the context of habitual actions. These classroom experiences, being more accessible to literate learners, help reinforce the acquisition of functional elements. If we consider Schwartz's (1993) concept of 'learned linguistic knowledge', we can infer that literacy likely facilitates the learning (though not necessarily the linguistic competence or acquisition) of morphosyntax. Schwartz distinguishes between 'learned linguistic knowledge' and linguistic competence. Learned linguistic knowledge results from explicit instruction and negative data, whereas linguistic competence develops through primary linguistic data (i.e., input). Explicit instruction and negative data do not contribute to linguistic competence. In other words, 'learned linguistic knowledge' and linguistic competence are two distinct processes. Notably, Schwartz's distinction aligns with Krashen's (1985) theory of 'learning vs. acquisition', where acquisition occurs through natural interaction with the language in meaningful communication, stimulating developmental processes similar to those in L1 acquisition. In contrast, learning occurs through classroom experiences that require learners to focus on form and systematically study the rules of the target language. Like Schwartz, Krashen argues that learning and acquisition are separate and cannot be integrated.

Vainikka and Young-Scholten's (1998) idea of triggers for the development of morphosyntax also presumes an indirect effect. In explaining the difficulty of acquiring functional morphology by adult L2 learners, Vainikka and Young-Scholten attribute the difficulty of bound morphemes, which results from phonology to the slow development of morphosyntax. For example, bound morphemes consist of units smaller than a phonological foot, making them difficult to perceive in the input. This difficulty is exacerbated for lower-literacy learners, who may lack the typical development of phonological awareness skills that are usually reinforced by literacy skills.

The delay in the development of morphosyntax associated with literacy suggests a specific role of literacy in the use of multi-word strings as overgeneralized forms. Although low levels of literacy contribute to this pattern, literacy distinctly reinforces functional elements, which are less salient in the input. As previously discussed, VP-stage learners in this study use multi-word strings as overgeneralized forms, unlike higher-stage learners who use only single function words. I have suggested that learners select forms based on the availability of functional elements in their OG stages. If that stage lacks functional elements (e.g., the bare VP stage), they use other elements available in the input.

While multi-word strings (e.g., 'you have' and 'in the') were used before verbs as overgeneralized forms, it is worth noting that these forms differ from fixed chunks (e.g., 'don't', 'doesn't', or the string 'I don't know'). These strings are not fixed or formulaic; rather, they appear to be formed from frequently occurring linguistic elements in the input and are used by these

learners regardless of their grammatical functions. In contrast, chunks are unanalyzed units stored in the mental lexicon and retrieved as wholes. However, these strings share with chunks that they are both salient and noticeable in the input. Lower-literate learners are more likely to notice and use these salient elements rather than less salient ones.

Therefore, I argue that these forms are not only available in the input but also salient and noticeable. These salient and noticeable elements are perceived more prominently than other IP-related elements (e.g., the copula 'is'). Under the usage-based approach, formulaic language is orthographically, auditorily, and perceptually more salient than morphosyntactic elements. It is more likely that salient forms will be perceived, attended to, and processed successfully (Wulff, 2019). Therefore, perceptual salience may explain 'success and failure' in learning forms through the approach of Focus on Form (Long, 2015, p. 60). According to Loewen and Sato (2018), the saliency of linguistic elements is indeed a mediating factor in L2 development during communicative interactions. Furthermore, the communicative value of formulaic language might be another possible reason for using multi-word strings as placeholders. Formulaic language, as opposed to morphosyntactic structures (e.g., the third-person singular -s) that are redundant in message comprehension, is vital for communicating meaning (Wulff, 2019). The communicative value of linguistic elements has been tied to their developmental patterns in L2 learning (Loewen & Sato, 2018). According to VanPatten's (2007) 'lexical preference principle', L2 learners analyze lexical items for meaning prior to morphosyntactic forms.

Another plausible explanation for using multi-word strings as placeholders might be that formulaic language is more noticeable. The noticeability of forms is influenced by the type of linguistic purpose (Loewen & Sato, 2018). Stimulated recall data revealed that corrective feedback for lexical errors was more consistently noticed than that for morphosyntactic errors (Mackey et al., 2000). Similarly, using a dual-mode system that differentiates between a memory-driven exemplar-based system and a rule-based analytic system, Yang and Lyster (2010) observed that rote-learned items were more noticeable than rule-based ones. Their findings demonstrated that corrective feedback focused on rote-learned items is more effective, regardless of the type of corrective feedback. Previous research has demonstrated that vocabulary items are more prominent than morphosyntax (Loewen & Sato, 2018). This research contributes to our understanding of how formulaic language, as larger units of lexical items, influences learners' degree of noticeability and is used as placeholders for morphosyntactic elements.

The salience and noticeability of forms as overgeneralized forms have also been suggested to explain why learners with low levels of morphosyntax or low literacy levels use them rather than functional elements such as the copula 'be'. Mocciaro (2019a) suggests that learners with low levels of morphosyntax use light verbs (e.g., 'to go') rather than functional elements such as the copula 'be' because these verbs are more salient in the input.

Conclusion

This study provides deeper insight into the role of literacy in the development of morphosyntax and the selection of overgeneralized forms. The findings suggest that low literacy may slow metalinguistic awareness or negatively impact the effectiveness of the triggers for morphosyntactic development, subsequently impeding this development and affecting the selection process for overgeneralized forms. However, before drawing definitive conclusions, researchers should conduct longitudinal studies that consider the relationship between metalinguistic awareness, literacy, and learners' stages of development, as lower-literate/lower OG

stage learners tend to use overgeneralized forms that are more salient in the input. This study also underscores the importance of fostering literacy skills in classrooms before introducing morphosyntactic aspects, as literacy may enhance metalinguistic awareness or the feasibility of less-salient morphosyntactic elements that facilitate morphosyntactic development. Educators should recognize that overgeneralization by lower-literate learners, like overgeneralization in L1 acquisition, is part of language development, particularly when considering the interaction between Universal Grammar and the input—specifically, the subconscious search for syntactic heads in the input, as adopted under Organic Grammar. While higher-literate/higher OG stage learners still use overgeneralized forms, these differ from those of lower-literate learners, suggesting that literacy skills do not prevent the occurrence of overgeneralization. However, literacy skills, along with metalinguistic information, may help learners consciously develop their morphosyntax, leading them to select IP-related elements as overgeneralized forms, as higher-literate or higher OG stage learners tend to do, or to avoid overgeneralizations altogether.

References

- Ehri, L. (1994). Development of the ability to read words. In R. Ruddell, M. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 323–358). International Reading Association.
- Hawkins, R. (2001). *Second language syntax: A generative introduction*. Wiley Blackwell.
- Julien, M., Van Hout, R., & Van de Craats, I. (2016). Meaning and function of dummy auxiliaries in adult acquisition of Dutch as an additional language. *Second Language Research*, 32(1), 49–73.
- Krashen, S. (1985). *The input hypothesis*. Longman.
- Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, 51(3), 285–329.
- Long, M. (2015). *Second language acquisition and task-based language teaching*. Wiley & Sons.
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22(4), 471–497.
- Mocciaro, E. (2019a). Emerging constructions in the L2 Italian spoken by low literate migrants. *Languages*, 4(1), 1–24.
- Mocciaro, E. (2019b). *The acquisition of L2 Italian morphosyntax by low- or non-literate adult learners* [Unpublished doctoral dissertation]. University of Palermo.
- Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, 15(2), 147–163.
- Tarone, E., Bigelow, M., & Hansen, K. (2009). *Literacy and second language oracy*. Oxford University Press.
- Vainikka, A., & Young-Scholten, M. (1998). Morphosyntactic triggers in adult SLA. In M.-L. Beck (Ed.), *Morphology and its interfaces in second language knowledge* (pp. 89–114). John Benjamins.
- Vainikka, A., & Young-Scholten, M. (2007). The role of literacy in the development of L2 morphosyntax from an Organic Grammar perspective. In N. Faux (Ed.), *Research, practice and policy for low-educated second language and literacy acquisition for adults: Proceedings of the 2006 LESLLA Conference* (pp. 123–184). Literacy Institute at Virginia Commonwealth University.

- Vainikka, A., Young-Scholten, M., Ijuin, C., & Jarad, S. (2017). Literacy in the development of L2 English morphosyntax. In M. Sosinski (Ed.), *Language and literacy teaching LESLLA students* (pp. 239–250). University of Granada.
- Vainikka, A., & Young-Scholten, M. (2019). Finding their heads: How immigrant adults posit L2 functional projections. In D. Arteaga (Ed.), *L2 grammatical representation and processing: Theory and practice* (pp. 116–138). Multilingual Matters.
- Van de Craats, I., & Van Hout, R. (2010). Dummy auxiliaries in the second language acquisition of Moroccan learners of Dutch: Form and function. *Second Language Research*, 26(4), 473–500.
- VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (pp. 115–135). Routledge.
- Wulff, S. (2019). Acquisition of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 19–37). Routledge.
- Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, 32(2), 235–263.
- Young-Scholten, M., & Strom, N. (2005). First-time L2 readers: Is there a critical period? In I. van de Craats, J. Kurvers, & M. Young-Scholten (Eds.), *Low-educated second language and literacy acquisition* (pp. 45–68). Landelijke Onderzoekschool Taalwetenschap (LOT).