

LESLLA Symposium Proceedings



Recommended citation of this article

Van de Craats, I. (2011). A LESLLA Corpus: L1 Obstacles in the Learning of L2 Morphosyntax. *LESLLA Symposium Proceedings*, 6(1), 33–48.
<https://doi.org/10.5281/zenodo.8004238>

Citation for LESLLA Symposium Proceedings

This article is part of a collection of articles based on presentations from the 2010 Symposium held at University of Cologne in Cologne, Germany. Please note that the year of publication is often different than the year the symposium was held. We recommend the following citation when referencing the edited collection.

Schöneberger, C., van de Craats, I., & Kurvers, J. (Eds.) (2011). Low-educated adult second language and literacy Acquisition (LESLLA): Proceedings of the 6th symposium. Centre for Language Studies.
<https://lesllasp.journals.publicknowledgeproject.org/index.php/lesllasp/issue/view/471>

About the Organization

LESLLA aims to support adults who are learning to read and write for the first time in their lives in a new language. We promote, on a worldwide, multidisciplinary basis, the sharing of research findings, effective pedagogical practices, and information on policy.

LESLLA Symposium Proceedings

<https://lesllasp.journals.publicknowledgeproject.org>

Website

<https://www.leslla.org/>

A LESLLA CORPUS: L1 OBSTACLES IN THE LEARNING OF L2 MORPHOSYNTAX

Ineke van de Craats, Radboud University Nijmegen

1 Why a LESLLA corpus?

It is only recently that second language (L2) acquisition researchers have started to focus their attention on adults with a low level of education and literacy rather than on highly educated and academic learners. Since the late nineties of the last century it became increasingly probable that illiterate individuals process oral language differently from literate adults. Illiterate adults turned out to have more problems repeating lists of pseudo-words than literates, or doing tasks with phoneme deletions and syllable reversals (e.g., Adrian, Alegria & Morais, 1995; Reis & Castro-Caldas, 1997). Neurological research by means of PET scans has even shown that there is more and different brain activation during pseudo-word repetition in literate than in illiterate participants. Learning to read and write an alphabetic script alters neural structures in the brain (Pettersson et al., 2000). For literates, the visual-graphic representation of a word is so closely linked to the phonological representation that they can play with the (written) symbols, without considering any semantic meaning. For fully illiterates, however, this manipulating of words without considering meaning is very difficult.

Results of other studies illustrate how much illiterates rely on semantic meaning because they cannot use visual-graphic strategies. Kurvers (2002) and Kurvers et al. (2006; 2007), for instance, concluded that illiterates did not view abstract and function words as words and had little metalinguistic and strategic skills. Tarone et al. (2007) found that illiterates could not process oral corrective feedback as easily as literates because of lacking the literacy skills allowing them to visually represent and compare their own utterance with that of the recast (Bigelow et al., 2006).

These findings are not only essential points to be aware of in language pedagogy and teaching of illiterate and low-literate learners, they are also of crucial importance for the interpretation of all other research on second language acquisition (SLA) that claims to identify universal cognitive processes involved in SLA. This research is almost exclusively based on tests and experiments carried out on academic students, often foreign-language students. One should wonder, as Tarone and colleagues do on page 1 of their 2009 book, whether an SLA theory of universal cognitive processes can be based exclusively on data from literate learners. As teachers of L2 literacy students often say that teaching this group is so different - and there are many signals in

research that this may be true - collecting longitudinal spontaneous and experimental data of LESLLA learners, particularly of those with less than two years of primary school education seems to be of high relevance for progress in this field.

It is not that LESLLA learners are completely absent from L2 research. More or less by accident, learners with a low level of education became involved in L2 research in some well-known longitudinal studies (e.g., Cancino et al., 1978; the ZISA project, Clahsen et al., 1983; the ESF project, Klein & Perdue, 1992) and the cross-sectional Lexlern study (Clahsen et al., 1991), because those studies aimed at observing to what extent adult learners were able to acquire a new language solely on the basis of aural input. For theoretical and practical reasons, adults with no other language knowledge than that of their mother tongue were the best subjects and those happened to be adults with little schooling. Their literacy level was not documented, as literacy and awareness of linguistic knowledge were not viewed as a contributing factor to L2 learning in those studies.

It is exactly in this respect that the interest of present days' LESLLA researchers differs and in which a LESLLA corpus would differ. The focus in a LESLLA study is on how a non-literate or low-literate learner copes with his restricted learning experience, in a tutored or untutored language or literacy learning context. A LESLLA researcher wants to know what is characteristic of those learners: is it a low pace of learning because they cannot (yet) read or the impact of another script system? Or do they rely more heavily on their L1 because they lack abstract knowledge of grammar and meta-linguistic skills? Is it stagnation or fossilization at an early stage?

Since the end of the last century, data of LESLLA learners have been collected not by accident, but by design. Kurvers & Van der Zouw (1990), and Kurvers (2002) collected data of adult L2 literacy learners in class. The former study focused on the development of reading in first time L2 readers, the latter on knowledge of language and script of illiterate L2 learners of Dutch. The Minneapolis Somali literacy study (Tarone et al. 2007; 2009) deals with illiterate and low-literate Somali learners of English and investigates the question as to what the impact of literacy on oral L2 use is. This study focused on three different issues: the impact of literacy on corrective feedback, on elicited imitation and on oral narratives. The fourth study is Strube's (2009) ongoing observation study of six L2 literacy classes in the Netherlands. She describes the learner's oral development, the teacher's feedback strategies and the learner's response. What would turn these four studies into corpora accessible to other researchers would be a digitalized speech recording with transcriptions and annotations (ideally, accompanied by a rough translation in English) on DVD or in a data bank which can be consulted on request.

The aim of the present contribution is to show, firstly, what such a corpus may look like by providing examples of what might be specific for the group of low-educated learners, and, secondly, that existing corpora should be made accessible and new corpora should be collected to enable comparison.

2 A LESLLA corpus

The data presented here and at the LESLLA symposium in Cologne meet most of the above criteria. All speech tasks were registered on a Sony mini-disc recorder, were

digitalized and converted into PRAAT ('talk' - a phonological transcription program, available online; Boersma & Weenink, 2003) and consecutively orthographically transcribed in Dutch; approximately half of the data have also been converted into CHAT, the transcription system related to the Chiles software. The PRAAT sound files with transcriptions are available on dvd for each participant and for each task in which spoken language is elicited.¹ An example of an utterance in PRAAT is given in Figure 1. In a PRAAT file different tiers can be used. The speech signal is in the main tier. The transcriber can select a part of the signal and listen to it in more detail so that a more precise transcription can be achieved. The second tier can be used for orthographic transcription. It is also very helpful that the intonation contour (the tier at the bottom of Figure 1) can be shown, which can often help in deciding where the utterance ends. New tiers can be added, e.g., for phonological transcription, for the transcription of what is said by an interlocutor, or for a translation in English (i.e., He goes to the window/the window), which has not been added in Figure 1, but would extend the accessibility. In the comment tier it is then explained that the self-correction (/ /) relates to the choice of the article. The rising contour is typical for this learner and probably for more Moroccan learners. A slight disadvantage is that the orthographic tiers can not directly be converted into a CHAT file.

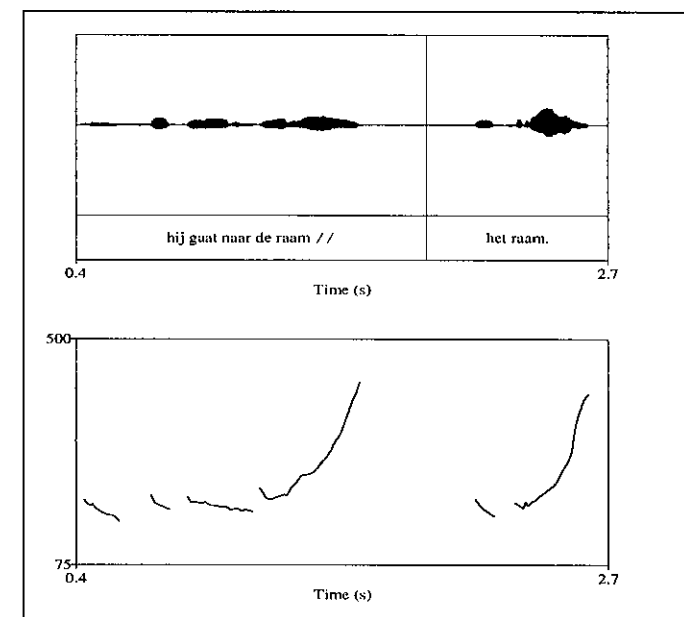


Figure 1: Sentence in picture-telling task uttered by the Moroccan subject Zobra in Cycle I

All participants had a low level of education or no formal schooling at all before arriving in the Netherlands. Some of them had attended a literacy class before they

¹ They can be obtained by sending an email to inekevandecraats@casema.nl

entered a 'regular' D(utch)SL course. Therefore, it was decided to call this corpus the LESLLA corpus. In what follows, first, the corpus will be described in this section. Sections 3-5 deal with aspects that might be characteristic for this group of learners, i.e., reading skills (Section 3), which are expected to be low, transfer from the L1 (Section 4), because these learners have low metalinguistic skills and no or little knowledge of other languages, and the realisation of verb morphology in narratives (Section 5), because it was suggested by Tarone, Bigelow & Hansen (2009: 97) that there might be a relationship between alphabetic print literacy level and the realisation of morphosyntactic features in oral narratives of low-literates.

2.1 Research design

The original aim of the study was to investigate where and when obstacles in the learning of L2 morphosyntax appeared and to what extent knowledge of the first language can explain these stagnations (i.e. a temporary or remaining stabilization) in tutored L2 acquisition. The study was set up as a longitudinal study in which the participants were observed for approximately 15-18 months, in three consecutive cycles of 5-6 months with three sessions each, nine sessions in all. In each cycle, the same tasks were administered ranging from free tasks (film-retellings, picture story-telling) to more controlled (e.g., a sentence completion task, a drag and drop task) and strictly controlled tasks (e.g., a sentence imitation task, a reading task, and a self-paced reading task). All tasks except the two reading tasks were designed to provide insight into the learner's reliance on the L1 (or L2) morphosyntactic structure of noun phrases, verb phrases and sentences. In the free tasks, the learner could freely produce L1- or L2-based structures, but in the tests the learner was really challenged by the design of the task (see for instance Figure 2). The repetitive character enabled us to compare the three cycles and to register even slight progress. The chance that participants would remember parts of the preceding cycle cannot be excluded, but would not help much because no corrective feedback was given by the researcher.

2.2 Participants

There were fifteen participants, all women, eight of them from Turkey and seven from Morocco. All participants had received little education in their native country and were learning Dutch in the instructional environment of a center for adult education. The teaching method can be best characterized as reflecting a communicative approach. Some participants also profited from contact with Dutch speaking neighbors, other mothers, and authorities. Since stagnation and its potential cause was the focus of research, at least half of the participants were judged by their teachers as having a stagnating learning process or running the risk to stagnate. As the impact of the mother tongue was seen as the most important factor for stagnation in beginning learners, speakers of two very different languages were chosen: Turkish and Moroccan Arabic. The fact that also the alphabetic writing systems of the two languages differ, was an additional aspect of this decision.

Table 1: Learner profiles of the Turkish participants at the start of data collection

Participant	Age	Years of schooling		Years in the Netherlands	Stagnation observed or expected by teacher
		Turkey	Netherl.		
Zilfi	30	5	1.5	11	no
Hülya	19	5	0.7	0.5	no
Emine	28	5	0.8	13	no
Hilal	19	5	1.8	2	yes
Ayfer	37	5	0.8	18	yes
Nazife	31	5	0.6	1	yes
Hatice	45	5	0.6	26	yes
Özlem	31	6	2.0	5	yes
Mean	30	5	1.0	9.5	

Table 1 shows that the Turkish participants were homogeneous in the years of schooling in Turkey; only Özlem had had some further education. Their age at the start of data collection varied from 19-45 (mean age 30). They were all spouses of so-called guest workers and arrived between 0.5 and 26 years (mean 9.5) ago.² They had all mastered a basic vocabulary after 0.7-2 years (mean: 1 year) of schooling in the Netherlands. Zilfi and Emine were able to communicate rather well, but communication with Nazife, Ayfer and Özlem was problematic.

At the end of the project, Hülya and Zilfi had attained level A1 of the Common European Framework or CEF (Council of Europe, 2001)³, Emine only for oral skills and Hilal only for writing; the other participants were below A1 for all four skills.

Table 2: Learner profiles of the Moroccan participants at the start of data collection

Participant	Age	Years of schooling		Years in the Netherlands	Literacy course Roman script	Stagnation observed/expected
		Morocco	Netherl.			
Mina	23	0	2.0	4	yes	no
Zohra	41	5	0.7	8	no	no
Soad	34	4	0.8	12	no	no
Najat	25	4	1.6	4	yes	yes
Hayat	22	5	2.0	2	yes	yes
Nezha	38	0	1.3	3	yes	yes
Fatima	27	7	1.8	5	no	yes
Mean	30	3.6	1.3	5.4		

The Moroccan learners (aged 22-41; mean 30), presented in Table 2, were all beginners as well (below level A1). They had been living in the Netherlands for 0-11 years (mean 5.4 years) before they started with the course. Mina, Zohra, Soad and Najat were able to communicate rather well, communication with Nezha and Fatima was problematic. Before taking this DSL course, four Moroccan participants attended a literacy class in

² Length of residence is not indicative for language contact because Muslim women often live(d) in the Netherlands with hardly any language contact with speakers of Dutch.

³ Level A1 is a very basic level characterized as Breakthrough. The basic vocabulary consisted of frequent and relevant words that occurred in the first five lessons of the textbook.

which they learned the Roman script; they were not illiterate in the Arabic script, although two of them, Mina and Nezha, had not attended elementary school. They learned reading from relatives at age 11 (Mina) and 20 (Nezha). At the beginning of the project all seven Moroccan learners could read a text in Arabic script and answer some simple comprehension questions.

At the end of data collection, Mina and Zohra had attained proficiency CEF level A2 ('Waystage'), Najat A1 only for speaking skills and the other participants were below A1 for all four skills.⁴

3 Development of reading skills in DSL course

Although it was out of the direct scope of the project, the reading skills of the 15 participants were assessed at the beginning of each cycle, by means of a short reading comprehension task in Dutch and a self-paced reading task. The reading tasks were administered because low reading proficiency or low processing speed might explain low scores on tasks aimed at assessing morphosyntactic knowledge. In the reading task the participant was asked to read a short text in Dutch and to answer one or two questions related to the text. She pushed the button when she stopped reading. The questions were intended to prevent the participants from rushing. In Table 3, the reading times in seconds are given for each cycle. The learners are roughly ranked in order of oral proficiency from top to bottom, partly based on CEF levels and partly on the results of the experiments.

Table 3: Reading pace in seconds for an L2 text

Turkish participants	Cycle			Mean	Moroccan participants	Cycle			Mean
	I	II	III			I	II	III	
Zilfi	131	108	111	117	Mina**	148	144	129	140
Hülya	86	110	100	99	Zohra	141	—	138	139
Emine	116	133	137	129	Soad	127	152	119	133
Hilal	143	103	132	126	Najat*	172	174	165	170
Ayfer	122	194	140	152	Hayat*	217	244	174	212
Nazife	117	98	99	105	Nezha**	307	264	225	265
Hatice	145	183	137	155	Fatima	120	123	161	135
Özlem	182	107	90	126					
Mean	130	129	118	126	Mean	176	183	158	172
Range				86-194					119-307

** = no reading instruction in L1 * = only reading instruction in Arabic script

When comparing the reading scores of the Turkish and Moroccan participants, the overall picture that can be derived from Table 3 is that the Turkish learners profit from their experience with the Roman script throughout the entire data collection. The ratio between the scores of the Turkish and Moroccan learners is 2:3. The non-stagnating learners are faster than the stagnating readers, apart from the Turkish learner Nazife –

⁴ There are specific national achievement tests and oral assessments (even one geared to low-educated learners) that are calibrated on communicative tasks of a level described in the CEF.

who said she loved reading – the Moroccan learner Fatima and the Turkish learner Özlem, who had more schooling and probably more reading experience than the other women.

In the self-paced reading task the participant was asked to read the sentence aloud, push the button, and recall the last word.⁵ Then the next sentence was presented. After two (or more) sentences, the participant had to recall the last words of each sentence in the order of presentation. In this way 3 sets of 2 sentences, 3 sets of 3 sentences and 3 sets of 4 sentences were presented. The computer registered the reading time of each sentence. Sentences of 16 syllables were presented. Table 4 shows the results for the learners in Cycle 1.

Table 4: Reading speed for sentences of 16 syllables on a self-paced reading task (Cycle I)

Task	Turkish participants	Moroccan participants
Mean time per sentence		
3 sets of 2 sentences	10 sec.	18 sec.
3 sets of 3 sentences	10 sec.	16 sec.
3 sets of 4 sentences	10 sec.	16 sec.

The ratio between the scores (i.e. speed) of the Turkish and Moroccan learners on this task is again 2 : 3. So, we can say that reading in another script affects the overall scores, but the learners differ much. Of the four learners who had attended a literacy class Nezha and Hayat progressed most in reading pace in the course of the project and Mina already read as fast as the literate Zohra.

4 L1 transfer

4.1 L1 transfer in noun phrases

The third task that we consider is a semi-controlled drag-and-drop task designed to provoke transfer from the L1. The learner knew that there were more blocks available to drag and drop than required for making a sentence.⁶ The task was designed in such way that both Turkish and Moroccan learners could construct a sentence corresponding to their L1 grammar. An example can illustrate the task. The word order of a possessive noun phrase in Turkish is: first the possessor, then the possessed

⁵ A Reading Span Test (Daneman & Carpenter, 1980) is traditionally used to measure working memory. We were interested in the participants' working memory capacity as a large working memory capacity may influence language learning (e.g., Baddeley, 2003). This experiment did not work out well for our participants, since the best learners seemed to fully process the sentences, whereas the other learners tended to neglect the processing of the sentence, but instead concentrated on recalling the last word. So, the more advanced learners got a longer reading time than the least advanced ones.

⁶ A reviewer remarked that this is a difficult task because of the use of the mouse and because of its abstract character. There were two items to try out the mouse and to grasp the purpose of the task. All participants already got some experience with the mouse in a preceding (literacy) course and a drag-and-drop task was not entirely new.

element, as in (1a). For Moroccan Arabic, it is just the opposite order (1b), whereas in Dutch both orders are allowed, as in (1c) and (1d).

- (1) a Hasan-in araba-si (Turkish)
 Hasan-3SG car -3SG
 'Hasan's car'
 b t-tumubil dya Hasan (Moroccan Arabic)
 the-car of Hasan
 c Hasan s/z'n auto (Dutch)
 Hasan his car
 d de auto van Hasan (Dutch)
 the car of Hasan

One of the items in this task was the one in Figure 2. Both the reaction time and the number of moves were registered, as the learner can try as many moves as she wants. Note that only type (1c), *Hasan z'n auto*, is correct and that three moves are sufficient to get the correct sentence. Note that type (1d) is not a possible answer in the item in Figure 2 because the analytic construction needs a(n) (un)definite article before the possessed element (DE auto).

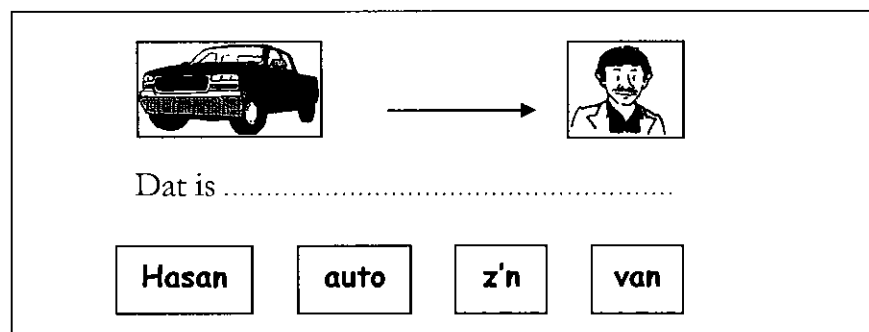


Figure 2: Item from the drag-and-drop task (target: 'Hasan z'n auto', Engl. 'Hasan's car')

This type of construction (1c) was difficult for both language groups, as can be derived from Table 5, with no correct answers in Cycle II for the Turkish participants, and no correct answers in Cycle I for the Moroccan learners. The percentage of correct answers was very low for both groups in the remaining cycles as well.

75% of the incorrect answers provided by the Turkish women were possessor-initial noun phrases (*Hasan van auto*; Hasan of car; 'Hasan's car') in Cycle I, 50% in Cycle II, and 37.5 in Cycle III. All these incorrect answers are based on the assumption that the preposition *van* 'of' is the genitive case of the possessor and not the preposition preceding the possessee NP. On the basis of their L1, the Turkish learners interpret the L2 differently from what native speakers and the Moroccan learners do. However, the answers of the Moroccan learners were all possessor-final noun phrases (*auto van Hasan*; car of Hasan; 'Hasan's car'), which are incorrect because a preceding article is required.

So, *van* is seen as positive evidence for transfer by most Turkish learners, and the absence of the article *de* is not seen as negative evidence by most Moroccan learners.

Table 5: Results of two items on word order in the noun phrase at the drag-and-drop task

	Turkish participants			Moroccan participants		
	% correct	moves	RT (sec.)	% correct	moves	RT (sec.)
<i>Hasan z'n auto</i> (1c) 'Hasan's car'		(3)				
Cycle I	12.5 %	4.1	27	0 %	4.4	35
Cycle II	0 %	6.4	41	28 %	3.5	19
Cycle III	12.5 %	7.7	32	17 %	3.5	21
<i>De opa van Bas</i> (1d) 'Bas' grandpa'						
Cycle I	25 %	5.1	25	100 %	3.8	34
Cycle II	25 %	6.2	27	100 %	3.8	26
Cycle III	25 %	8.3	43	100 %	4.3	21

The second item in Table 5, *de opa van Bas*, the grandpa of Bas ('Bas' grandfather'), was a type (1d) construction in which the possessee is preceded by an article, completely corresponding to the Moroccan Arabic construction type (1b), resulting in a 100% correct score for the Moroccan learners.

These results also show that the Turkish learners struggle much harder but get worse results, i.e., fewer correct scores with more moves. The assumed reason is that they are restructuring their L1 grammar (they reanalyze the category of the case marker *van* 'of' into a preposition), whereas the Moroccans simply transfer their L1 structure.

4.2 L1 transfer in verb phrases

Similar items in which the learner is elicited to follow her L1 grammar, were constructed for verb phrases in the drag-and-drop task. Just as was the case for noun phrases, Dutch allows two syntactic positions for the lexical verb: in sentence-final position when the verb is non-finite (2b) and in second position when it is finite (Dutch is a Verb Second language; see e.g. Den Besten, 1989), as in (2a). In the latter case, the object precedes the non-finite verb (OV order). Turkish has the same basic word order, but then the lexical verb is finite (2c), while Moroccan Arabic has a SVO (2d) or VSO order.

- (2) a Bas koop-t een boek SVfinO (Dutch)
 Bas buy- 3SG a book
 'Bas buys a book.'
 b Bas moet een boek kop-en SAuxOV (Dutch)
 Bas must.SG a book buy-NONFIN
 'Bas must buy a book.'
 c Ahmet kitap al-iyor SOVfin (Turkish)
 Ahmet book buy-3SG
 'Ahmet buys a book.'

- d Abder kayehder l-Eerbiya SVfinO (Moroccan Arabic)
 Abder speak.3SG the-Arabic
 'Abder speaks Arabic.'

A Dutch main sentence may, therefore, provide evidence for both OV and VO order, but OV order is restricted to non-finite and VO to finite contexts. The question arises whether, and if so, how long the present learners of Dutch are guided by their L1 word order. The results on two relevant items of the drag-and-drop task provide some indication.

In the first item ('Bas buys stamps') three blocks could be dragged and the learner could choose between a finite (*koopt*) and an infinite form (*kopen*), which could be placed in the second position (3a) or (3d) or at the end of the sentence, as in (3b) and (3c). The target sentence (3a) can be attained in minimally two moves.

- (3) Stimulus: prompt + 3 blocks = Bas /postzegels/kopen/koopt/
 a Bas ... *koopt* postzegels. (Vfin – position and form correct)
 Bas buys stamps
 b Bas ... postzegels *koopt*. (Vfin – position not correct)
 c Bas ... postzegels *kopen*. (Vfin – position and form not correct)
 d Bas ... *kopen* postzegels. (Vfin – form not correct)

The second item ('Freek gets a fine') given in (4) has a similar target sentence, the only difference being the order of presentation and the number of blocks, four in this case, as a prompt is lacking. The target sentence can be moved in minimally three moves.

- (4) Stimulus: 4 blocks = /krijgen/Freek/een bon/krijgt/
 Freek *krijgt* een bon.
 Freek gets a ticket.

The results for the two items are given in Table 6. For the Moroccan learners, there was a strong similarity with their L1 structure, which resulted in an almost 100% correct score: both the position and the morphological form were correctly chosen. The scores of the Turkish learners are around chance level (50% plus or minus 12.5) over the three cycles. For them, the relationship between the finite verb and the second position was not evident, as finite verbs basically occur in sentence-final position in Turkish. The error scores in Cycles I and II are all cases of a verb (finite or infinite) in sentence-final position, only in the last cycle two infinite forms in finite position occurred.

This lack of certainty becomes clearly manifest in (5), in which Zilfi produced both forms in two different syntactic positions: the two forms are in the correct position, but the nonfinite form is redundant.

- (5) Bas *koop-t* postzegels *kop-en*
 Bas buy-3SG stamps buy-NONFIN

Table 6: Results of two items on word order in the sentence at the drag-and-drop task

	Turkish participants			Moroccan participants		
	% correct	Moves (2)	RT (sec.)	% correct	Moves (2)	RT (sec.)
<i>Bas koopt postzegels</i> 'Bas buys stamps'						
Cycle I	62.5%	2.9	19	100%	4.1	36
Cycle II	50%	3.3	27	100%	2.4	17
Cycle III	37.5%	3	33	100%	2.8	23
<i>Freek krijgt een bon</i> 'Freek gets a ticket'			(3)			
Cycle I	50%	3.8	17	100%	3.1	35
Cycle II	62.5%	4.4	19	86%	3.6	27
Cycle III	62.5%	3.8	18	100%	3.8	20

To conclude it can be stated that the Moroccan learners – with less schooling and literacy – are more aware of variable, morphological features and the position of the verb and their relationship than the Turkish learners in this project. Can this be confirmed by the morphosyntactic development of the verb in a relatively free task such as a film retelling and a picture-telling story?

5 Morphosyntactic features and default forms in narratives

The LESLLA corpus contains three narratives per cycle. We scanned the two narratives with the highest number of words for all verb forms meant to refer to 3SG of the present tense, because subject-verb agreement can only reliably distinguished from the stem and the infinitive of a verb in this context, as can be seen in Table 7. Henceforth, verb forms consisting of stem and stem+t are denoted as short forms, and verb forms consisting of stem+en (often pronounced as -e in spoken Dutch) as long forms.

Table 7: Dutch inflectional paradigm for regular verbs in the present tense

Person + number	-suffix	Example: (<i>pakken</i> 'to take')		Word length in syllables	Morpho-syntactic
1sg	-Ø	ik pak	'I take'	Short	Finite
2sg	-t/-Ø	jij pakt/pak je?	'You take'	Short	Finite
3sg	-t	hij pakt	'He takes'	Short	Finite
1,2,3 pl	-en	wij/jullie/zij pakken	'We/you/they take'	Long	Finite
Infinitive	-en	pakken	'to take'	Long	Nonfinite

In Figure 3, all long forms and short forms of lexical verbs are put together for three Turkish and three Moroccan learners (one of them being the least proficient learner (at the top) with the most schooling, the other two the most advanced learners at the bottom with one of them – Mina – without any schooling in L1).

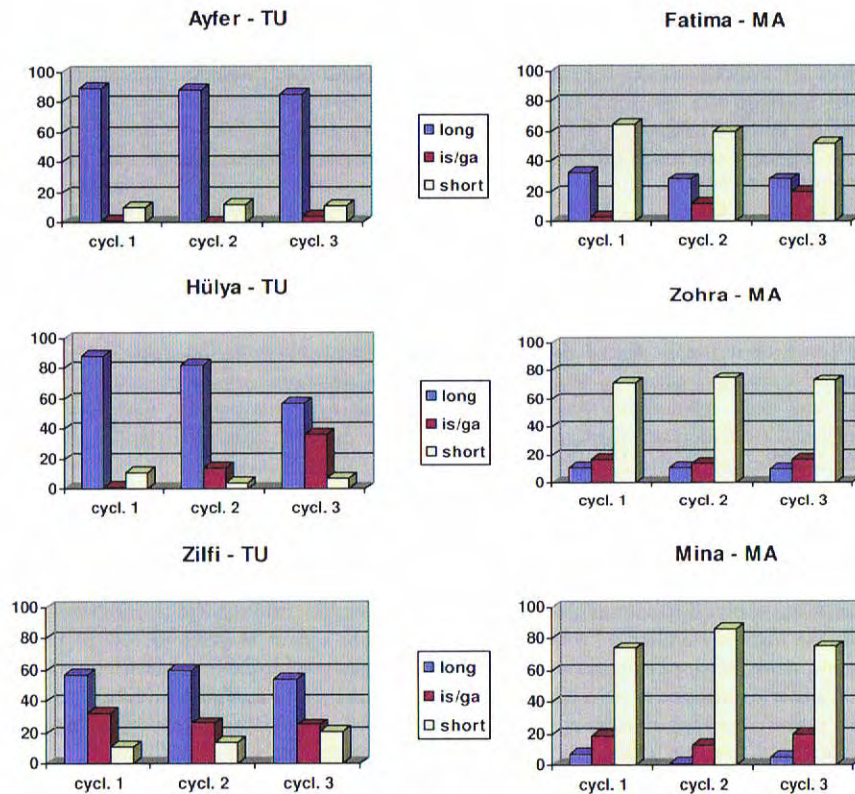


Figure 3: The morphosyntactic development of three Turkish and three Moroccan learners over three cycles of six months; *is* (be-3sg) and *ga/gaat* (go-1/3sg) are dummy auxiliaries discussed later.

The differences between the three Turkish and three Moroccan learners are similar to those between the learners that are not represented in Figure 3. The number of occurrences of long forms versus short forms in the two groups was analyzed by computing logit values (a value of .5 was added to all frequencies). An ANOVA was applied to the logit values of the total number of long versus short forms, with cycle (within-subjects factor) and L1 group (between-subjects factor) as independent variables. The result was a significant effect for L1 group ($F = 17.869$, $df = 1,12$, $p = .001$, partial $\eta^2 = .598$), as expected. The Turkish learners produced more long forms. No cycle effect was found ($F = 1.282$, $df = 2,24$, $p = .296$, partial $\eta^2 = .097$), but the interaction between cycle and group was significant ($F = 5.313$, $p = 2,24$, $p = .012$, partial $\eta^2 = .307$). Neither literacy, nor the assumed difficulty for illiterates and low-literates to process morphological markers on the verb (as hypothesized by Tarone et al. 2009, chapter 6) seem to be the cause of these huge differences, rather the interplay

between L1 and L2 seem to be the key factor. If this is the case, what makes Moroccan learners to seem - or to be - the more advanced learners of Dutch?

The cause may be that beginning learners take default verb forms instead of carrying out an agreement procedure. Turkish learners may take a different default form from the Moroccan learners depending on the interplay between L1 and L2. Previous research has shown that Turkish learners prefer a long form, and Moroccan individuals a short form.⁷ Usually the infinitive is considered the default form, but that is not necessarily the case at the beginning of the language learning process. Moroccan learners, for instance, may have various reasons for taking a short form: (i) there is no infinitive in Moroccan Arabic, the 3SG perfect has that function; (ii) the pronunciation of a schwa as an independent syllable – a Dutch infinitive ends on *-(e)n*, e.g. *kijk^{en}* 'to look' – at the end of a word is a deviation from the L1 norm and therefore a challenge; (iii) the position of the finite verb in the sentence is similar. Two examples of short forms used in a non-finite context are given in (6).

- (6) a kan niet [_{VP} gaa-t fiets] Najat, Moroccan learner
 can not go-3SG cycle.STEM
target: zij kan niet gaan fietsen
 'she cannot start cycling'
- b ik ga buiten [_{VP} speel-t] Fatima, Moroccan learner
 I go.STEM/1SG outside play-3SG
target: ik ga buiten spelen
 'I am going to play outside.'

Different reasons such as a similar syntactic position of the infinitive in Dutch and a finite verb in Turkish may lead the Turkish learner to prefer a long form. Whereas the Moroccan learner has to learn the form and the function of an infinitive, the Turkish learner has to learn that a finite verb occurs in the second position of a sentence (the verb has to be moved from the VP to a functional projection). This process starts with modals, but if there is no modal aspect, the learner may insert a dummy auxiliary, a verb form that carries only grammatical information, e.g. person and number, as shown in (7a), in which *is* (be-3SG) is inserted between the subject and the negator. This construction is not allowed in native Dutch.

- (7) a Vader *is* niet komen Zilfi, Turkish learner, Cycle I
 Father is not come.NONFIN
 'Father does not come.'
- b Vader *is* niet kom Cycle II
 Father is not come.STEM
- c kom niet die vader Cycle III
 come not that father

What makes the series in (7) interesting is that Zilfi utters this sentence at the same scene in the retelling with an interval of about five months, thus showing how the

⁷ This is exactly the reason why the terms short and long are preferred over finite and infinite: there is no clear relationship between finiteness and morphological marking at the beginning.

dummy auxiliary form *is* accompanies the movement of the lexical verb and takes over the person and number features of the finite verb (see for morphosyntactic details: Van de Craats, 2009) before the (entire) verb is moved to the position in the beginning of the sentence in (7c), in which agreement is still not correct. It seems simpler to express these features independently than linked to a lexical verb when movement of the entire verb is involved.⁸

A similar process occurred for the Moroccan learners, who preferred another dummy auxiliary *gaan* ('to go'), as in (8a) and (8b), in which *ga* (go.STEM) carries the person, number and tense features without the meaning it normally has. Note that in both sentences the learner reports an action that is already taking place and not one that is going to happen (see for more information: Van de Craats & Van Hout, 2010). In (8a), the dummy form is linked with a long form, in (8b) with a short form.

- (8) a Sneeuwman *ga* kijk-en tableaux Fatima, Moroccan learner
snowman go.STEM look-NONFIN paintings
'The snowman is looking at the paintings.'
- b dan *ga* loop naar raam Hayat, Moroccan learner
then go.STEM walk.STEM to window
'Then he is walking to the window.'

The number of dummy auxiliaries – either *is* or *ga(at)* – produced by three learners of each group is given in Figure 3, in which it can be observed that the more advanced Turkish learners Hülya and Zilfi used more dummy auxiliaries than Ayfer, who does not show development. For Hülya the long forms are disappearing in favour of dummy auxiliaries, for Zilfi the dummy auxiliaries are disappearing in favour of finite forms. For the Moroccan learners the picture is different. The least advanced learner, Fatima, produced more dummy auxiliaries over time, and fewer short forms (probably default forms), the more advanced Zohra and Mina still produced *ga*-forms, but with the target meaning of near future (not visible in the graph) and almost no long forms instead of finite forms.

To conclude it can be stated that dummy auxiliaries realise one of more features that are normally part of a lexical verb and may emerge when bound verb morphology has not fully been acquired yet. They disappear after a specific developmental stage, but can remain for quite a long time in the speech of vulnerable learners such as LESLLA learners and SLI children (e.g., Jolink, 2005; de Jong, 1999).

6 Conclusions

For all aspects of L2 acquisition we have been dealing with – i.e. reading pace, L1 transfer, defaults forms and dummy auxiliaries – we do not know how characteristic they are for the population of low-literate and low-schooled learners, because comparison with other groups is problematic for the simple reason that (either longitudinal or cross-sectional) data are lacking, including data of highly educated learners. Therefore we do not know whether insertion of dummy auxiliaries is related

⁸ Similar processes are found for L2 English, e.g., by Fleta (2003) and for L2 German by Haberzettl (2003).

to slow development or due to a lack of linguistic experience with languages in which inflectional features are overtly realised. Anyway, it is not caused by a lack of literacy alone since both literate Turkish and low-literate Moroccan learners use them. We do not know either if L1 transfer is more persistent in LESLLA-learners, but we do know that absence of correct morphological verb marking is not only characteristic for illiterates, but also for the low-literate and moderately literate learners in the present corpus. For Dutch, however, absence of verb marking manifests itself not as a bare verb (like the Somali learners in Tarone et al., 2009) but as a default form, either a long form (infinitive) or a short form (unanalysed finite form). Although Tarone et al. found more bare verbs in the low-literate than in moderately literate group, the differences within the groups were so large that there was no statistic significance for the relationship between literacy and use of verb morphology. Other corpora with LESLLA-learners may confirm this, but we also need cross-sectional or longitudinal data from highly literate learners to know if lack of inflectional morphology is not more typical for beginning learners than for illiterate learners or for highly educated learners with first languages lacking verb morphology like the highly educated Patty in Lardiere (1998).

References

- Adrian, J.A., Alegría, J. & Morais, J. (1995). Metaphonological abilities of Spanish illiterate adults. *International Journal of Psychology* 30: 329-353.
- Baddeley, A.D. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, 36, 189-208.
- Bigelow, M., delMas, B., Hansen, K., & Tarone, E. (2006). Literacy and the processing of oral recasts in SLA. *TESOL Quarterly* 40: 1-25.
- Boersma, P., & Weenink, D. (2003). PRAAT Phonological Transcription, version 4.2.34. Available online via <http://www.praat.org> (May 2010).
- Cancino, H., Rosanky, E., & Schumann, J. (1978). The acquisition of English negatives and interrogatives by native Spanish speakers. In E. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 207-230). Rowley, Massachusetts: Newbury House.
- Clahsen, H., Vainikka, A., & Young-Scholten, M. (1991). Lernbarkeitstheorie und Lexikalisches Lernen. *Linguistische Berichte*, 130: 466-477.
- Clahsen, H., Meisel, J., & Pienemann, M. (1983). *Deutsch als Zweitsprache. Der Spracherwerb ausländischer Arbeiter*. Tübingen: Gunter Narr.
- Council of Europe, 2001: A Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.
- Craats, I. van de (2009). The role of *is* in the acquisition of finiteness by adult Turkish learners of Dutch. *Studies in Second Language Acquisition*, 31, 1, 59-92.
- Craats, I. van de, & Hout, R. van (2010). Dummy auxiliaries in the second language acquisition of Moroccan learners of Dutch: Form and function. *Second Language Research*, 26: 473-500.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466.

- Den Besten, H. (1989). *Studies in West Germanic syntax*. Amsterdam: Rodopi.
- Fleta, M.T. (2003). Is-Insertion in L2 grammars of English: A step forward between developmental stages? In H. Zobl, & H. Goodluck (Eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)* (pp. 85-96). Somerville, MA.: Cascadia Proceedings Project.
- Haberzettl, S. (2003). 'Tinkering' with chunks: Form-oriented strategies and idiosyncratic utterance patterns without functional implications in the IL of Turkish speaking children learning German. In C. Dimroth, & M. Starren (Eds.), *Information structure and the dynamics of language acquisition* (pp. 45-63). Amsterdam: John Benjamins.
- Jolink, A. (2005). Finite linking in normally developing Dutch children and children with specific language impairment. *Zeitschrift für Literaturwissenschaft und Linguistik*, 35, 140: 61-80.
- Jong, J. de (1999). *Specific language impairment in Dutch: Inflectional morphology and argument structure*. RU Groningen, doctoral dissertation.
- Klein, W., & Perdue, C. (1992). *Utterance structure. Developing grammars again*. Amsterdam/Philadelphia: John Benjamins.
- Kurvers, J. (2002). *Met ongeletterde ogen. Kennis van taal en schrift van analfabeten*. [With illiterate eyes. Knowledge of language and script of illiterates.] Amsterdam: Aksant.
- Kurvers, J., Hout, R. van, & Vallen, T. (2006). Discovering features of language: metalinguistic awareness of adult illiterates. In I. van de Craats, J. Kurvers & M. Young-Scholten (Eds.), *Low-Educated Second Language and Literacy Acquisition. Research, Policy, and Practice. Proceedings of the Inaugural Symposium Tilburg 2005* (pp. 69-88). Utrecht: LOT.
- Kurvers, J., Hout, R. van, & Vallen, T. (2007). Literacy and word boundaries. In N. Faux (Ed.), *Low-Educated Second Language and Literacy Acquisition. Research, Policy, and Practice. Proceedings of the Second Annual Forum* (pp. 45-64). Richmond: The Literacy Institute.
- Kurvers, J. & Zouw, K. van der, (1990). *In de ban van het schrift. Over analfabetisme en alfabetisering in een tweede taal*. [Under the spell of script. On illiteracy and learning to read and write in a second language]. Amsterdam/Lisse: Swets & Zeitlinger.
- Lardiere, D. (1998). Dissociating syntax from morphology in a divergent end-state grammar. *Second Language Research* 14: 591-633.
- Pettersson, K.M., Reis, A., Askelof, S., Castro-Caldas, A., & Ingvar, M. (2000). Language processing modulated by literacy: A network analysis of verbal repetition in literate and illiterate subjects. *Journal of Cognitive Neuroscience*, 12: 364-382.
- Reis, A., & Castro-Caldas, A. (1997). Illiteracy: A cause for biased cognitive development. *Journal of the International Neuropsychological Society*, 3: 444-450.
- Strube, S. (2009). What do teachers do? A look at the oral skills practices in the LESLLA classroom. In I. van de Craats, & J. Kurvers (Eds.), *Low-Educated Second Language and Literacy Acquisition. Proceedings of the 4th Symposium* (pp. 49-62). Utrecht: LOT.
- Tarone, E., Bigelow, M., & Hansen, K. (2009). *Literacy and Second Language Oracy*. Oxford: Oxford University Press.
- Tarone, E., Bigelow, M., & Hansen, K. (2007). The impact of alphabetic print literacy level on oral second language acquisition. In N. Faux (Ed.), *Low-Educated Second Language and Literacy Acquisition. Research, Policy, and Practice. Proceedings of the Second Annual Forum* (pp. 99-122). Richmond: The Literacy Institute.