

LESLLA Symposium Proceedings



Recommended citation of this article

Mocciaro, E. (2021). The Acquisition of L2 Italian by LESLLA Learners: Collecting and Analysing Data. *LESLLA Symposium Proceedings*, 14(1), 267–291.

Citation for LESLLA Symposium Proceedings

This article is part of a collection of articles based on presentations from the 2018 Symposium held at University of Palermo, Palermo, Italy. 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.

D'Agostino, M., Mocciaro, E. (Eds.) (2021). *Languages and literacy in new migration. Research, practice and policy. Selected papers from the 14th Annual Meeting of LESLLA (Literacy education and second language learning for adults)*. Palermo University Press.
<https://lesllasp.journals.publicknowledgeproject.org/index.php/lesllasp/issue/view/478>

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/>

The acquisition of L2 Italian by LESLLA learners: collecting and organising data

Egle Mocciaro

In the last decades, Italian research has provided accurate descriptions of L2 Italian acquisition morphosyntax, but learners' home literacy has never been assumed as an explicit variable in data collection and analysis. However, limited literacy and/or schooling are important components of migrants' background in recent flows to Italy (and Europe) and deserve new attention. The research project "The acquisition of L2 Italian morphosyntax by low and non-literate learners" (University of Palermo, 2016-2019) is an effort in this direction. Part of the broader research and pedagogical activity carried out at the School of Italian for Foreigners of Palermo, the project aims at verifying the degree of applicability of existing descriptions of L2 Italian to learners with limited literacy and at comparing the resulting data with research products on other L2 acquired by this population of learners. In this paper, the theoretical and methodological assumptions on which the research was based are presented and discussed.

Keywords: L2 Italian morphosyntax, learners' limited literacy, adult learners.

1. LESLLA learners: reasons for a corpus

During the last three decades, Italian research has provided descriptions of L2 Italian acquisition paths and learners' morphosyntax. An important aspect of this research has been its corpus-based character, which has allowed the systematic analysis of a substantial amount of data (Chini 2021; Giacalone Ramat 2003)¹.

¹ Cf. the databank of the Pavia Project (see Andorno 2001). Other important corpora of L2 Italian which collect oral and/or written texts are: ADIL2 – Archivio Digitale di Italiano L2 (Università per Stranieri di Siena; see Palermo 2005, 2009), LIPS – Lessico Italiano Parlato di Stranieri (Università per Stranieri di Siena; see Gallina and Barni 2009), Corpus parlato di italiano L2 (Osservatorio sull'italiano di stranieri e sull'italiano parlato all'estero, Università per Stranieri di Perugia; see Atzori and Spina 2009), Italiano scritto L2 (Università di Salerno; see Turco and Voghera

However, none of these studies assumed learners' early literacy in a home language as an explicit variable in data collection and analysis. As a consequence, the role that this variable plays on the acquisition of L2 Italian has so far remained largely unexplored². Nonetheless, limited literacy is an important component of migrants' background, especially in recent flows towards Italy (and Europe in general) and requires new attention and in-depth analysis so that the relevance of this neglected factor can be verified on the basis of consistent data. The main context in which such research should be conducted is represented by the international forum LESLLA (Literacy Education and Second Language Learning for Adults, www.leslla.org).

The research project "The acquisition of L2 Italian morphosyntax by low and non-literate learners", started in 2016 at the University of Palermo and concluded in 2019, was an effort in this direction³. The project was part of the broader research activity on adult and young adult migrants carried out at the School of Italian language for Foreigners of Palermo (henceforth, ItaStra) since 2012. In this local context, an increasing amount of data on migrants' interlanguage has been collected over the years, largely resulting from interviews conducted for various preliminary studies (see Mocciaro 2019, 2021).

In 2016-2019, this nucleus has been expanded on by new data collection, specifically designed for the research project on the acquisition of morphosyntax. The new sub-corpus consists of the recordings and transcriptions of young adult migrants' speech, both literate and with limited literacy, in different stages of their acquisition path. This LESLLA learner corpus is the starting point for:

- verifying the degree to which the existing descriptions of L2 Italian also apply to LESLLA learners (in terms of route, rate and end-state of the second language acquisition process);

2010), VALICO – Varietà Apprendimento Lingua Italiana Corpus Online (Università di Torino; see Corino and Marellò 2017).

² To my knowledge, the only systematic study on the role of L1 literacy on the acquisition of L2 Italian is Maffia and De Meo (2015) on the development of L2 prosodic competence in low-literate Senegalese learners.

³ This paper is based on the presentation given at the 14th Annual LESLLA Symposium, held in Palermo in 2018. The research was completed in 2019 and the results are now published in Mocciaro (2020).

- analysing LESLLA learners’ interlanguages in the light of current theoretical insights on the role of literacy in second language acquisition;
- comparing the resulting data with the research products on other L2s acquired by such learner population⁴.

In this paper, I will present and discuss the theoretical background of the project (Section 2). The attention will be focused on the research context (Section 3.1), the participants’ profile (Section 3.2) and the criteria for data collection (Section 3.3) and elaboration (Section 3.4). This first phase of the research has brought to light some difficulties, representing, however, important challenges at the theoretical and methodological level (Session 4). Two of them appear to be specifically related to the migrant population, namely the difficulty of isolating the variable “literacy” and the high attrition rate in the sample that makes a longitudinal survey particularly difficult. In addition, some problematic aspects related to data elaboration criteria are addressed, which involve learner corpora in general. It is suggested that the choices made in this respect may affect the analysis of LESLLA learners’ interlanguage, eventually bringing to the fore possible preferences in selecting specific form/function pairs.

2. L1(s) literacy and L2 acquisition

2.1. Studies on literacy and Second Language Acquisition

The impact of L1(s) literacy is still a peripheral area in Second Language Acquisition (SLA) research and is largely carried out by scholars who gravitate to some extent towards the LESLLA forum. In this context, there is consensus that L1 (alphabetic) literacy correlates with certain aspects of L2 acquisition, although the nature of this relationship is anything but clear and different scholars emphasise different aspects, also depending on their theoretical premises.

⁴ Cf. for instance the Dutch LESLLA corpus described in van de Craats (2011) and Sanders et al. (2014).

Vainikka and Young-Scholten (2007a: 144) summarised the debate around this topic as in Table 1, which, after more than ten years, still illustrates well the main tendencies within the few LESLLA studies which have been carried out since then. Table 1 shows the contrast between:

- the strong generativist hypothesis (arguing against any external, not strictly linguistic influence on language acquisition) and the weak continuity generativist hypothesis (i.e., the Organic Grammar approach which admits the possible influence of external factors such as literacy on L2 acquisition, cf. Vainikka and Young-Scholten 1998, 2006, 2007b);
- the generativist perspective as a whole and the line of research directly inspired by cognitive psychology experimental studies on phonological awareness and working memory (cf. the research conducted by Tarone and colleagues since the early 2000s).

	HYPOTHESIS	TESTABLE BY	EVIDENCE FROM	HYPOTHESIS STATUS
STRONG GENERATIVIST HYPOTHESIS	Literacy does not affect acquisition	Looking at L2 learners regardless of their literacy, etc.	Existing studies of immigrants	supported
INDIRECT INFLUENCE HYPOTHESIS	Literacy affects morphology which in turn affects syntax	Comparing nonliterate and literate L2 learners	Tarone, Bigelow and colleagues' work	some support
INDIRECT INFLUENCE HYPOTHESIS II	Literacy affects phonology which affects operation of triggers (morphology) which affects syntax	Comparing nonliterate and literate L2 learners	Weak generativist hypothesis: Vainikka and Young-Scholten	some support
INTERFACE HYPOTHESIS	Literacy affects processing which affects acquisition of morphology and syntax	Comparing nonliterate and literate L2 learners	Tarone, Bigelow and colleagues	some support

Table 1. The role of literacy in the acquisition of L2 morphosyntax (Vainikka and Young-Scholten 2007a: 144).

The cognitive trend rests on the assumption that alphabetic literacy affects the ability to segment the speech into non-semantic units – in particular, into phonemes – and to manipulate these units (Castro-Caldas 2004; Goswami and Bryant 1990; Huettig and Mishra 2014; Reis and Castro-Caldas 1997). On the other hand, literacy does not affect the ability to process oral speech in semantic units. The awareness of phonology is an explicit, that is, metalinguistic skill, whereas the semantic processing is implicit. Illiterates and literates process oral language differently, as only alphabetic literacy provides strategies to process oral language segments irrespective of the semantic content. As Reis and Castro-Caldas (1997: 445) observe,

Learning to match graphemes and phonemes is learning an operation in which units of auditory verbal information heard in temporal sequence are matched to units of visual verbal information, which is spatially arranged. This type of treatment of auditory verbal information modulates a strategy in which a visual-graphic meaning is given to units that are smaller than words, and thus independent of their semantic representation. [...] If we, as normal adult readers, are asked to spell a word, we evoke a visual image of its written form. The awareness of phonology also allows us to play with written symbols (which can be transcoded to sounds) to form pseudo-plausible words, independently of semantics. Therefore, learning to read and write introduces into the system qualitatively new strategies for dealing with oral language; that is, conscious phonological processing, visual formal lexical representation, and all the associations that these strategies allow.

Phonological awareness also entails the ability to segment the oral input into words as phonological units, independent of lexical semantics, that is, to identify word boundaries in the speech continuum (Reis et al. 2007).

While the impact of literacy in explicit processing operations is widely recognised in LESLLA studies, there are different views on the role that the explicit processing in turn plays on SLA. Some assume that explicit processing of the oral input is a necessary condition for adults to acquire a second language, as postulated in Schmidt's (1990) *noticing hypothesis*. If a learner can't consciously notice formal segments in the speech stream, then s/he will not acquire relevant functional units, such as grammatical morphemes (e.g., English plural *-s*, third person singular *-s*, past tense *-ed*). As a

consequence, learners who have not experienced literacy will produce bare forms (i.e., not inflected verbs and nouns) more often than literates do (cf. Bigelow and Tarone 2004; Tarone 2010; Tarone and Bigelow 2005; Tarone, Bigelow and Hansen 2007, 2009). In addition, if non-literates lack a non-semantic notion of word (Kurvers, van Hout and Vallen 2007; Onderlinden, van de Craats and Kurvers 2009), we should also expect that they will struggle in acquiring purely functional words (e.g., articles, auxiliaries etc.)

Alternatively, other assume that SLA is a fully implicit process, activated through the pure exposure to the linguistic input (cf. Schwarts 1993). External factors and general cognitive mechanisms (such as metalinguistic awareness) cannot be conceived of as direct input for SLA. Along these lines, Young-Scholten and Strom (2006) find a positive correlation between morphosyntax development and literacy, but are cautious in hypothesising a relation of causality, as different factors can still interfere. According to Vainikka and Young-Scholten (2007a: 143), literacy affects phonological attainment, which in turn may result in incomplete analysis of morphological constituents. This analysis develops the hypothesis formulated in previous research (Vainikka and Young-Scholten 1998: 97) that whereas bound morphemes (e.g., inflectional affixes) typically trigger L1 morphosyntax acquisition, it is free morphemes that do so in L2 acquisition, and it is possible that a bound morpheme can never act as a trigger: "If a particular parameter can only be triggered by a bound morpheme, such a parameter may be difficult or impossible to set in L2 acquisition, resulting in a fossilized non-target grammar". They propose that the distinction between bound and free morphemes as triggers may depends on phonology (p. 106): "Free morphemes such as auxiliaries typically constitute at least a phonological foot, while bound morphemes typically involve units smaller than a foot. Lack of phonological attainment may in turn result in incomplete analysis of sub-foot constituents in the learner's L2". Vainikka et al. (2017) claim that non-literate learners tend to overgeneralise specific function words and sequences (*placeholders*) to mark morphosyntactic functions in L2 English (e.g., *in the* to mark progressive aspect, as in *in the drink*, *in the no cooking*); while these forms can be not directly related to the actual verbal head, their occurrence still shows that non-

literate learners are able to subconsciously identify functional forms in the input.

An aspect which clearly emerges from all the LESLLA studies here reported is the necessity – yet the difficulty – to isolate the different variables at work in adult migrants' acquisition process which correlate and may interact in a quite complex way:

- first, literacy and schooling/education should be measured separately, also because they do not necessarily implicate each other;
- second, literacy should be kept distinct from (the quantity and quality of) exposure to the target language, since low levels of literacy typically correlate with migrants' low interaction with native speakers, which results in a low amount of linguistic input.

As a consequence, the two variables easily overlap and it is a difficult task to assess which of them does actually affect morphosyntax acquisition. Not to mention that low literacy itself corresponds to as low exposure, as learners cannot access the written input.

2.2. Italian research on second language acquisition

The most systematic Italian SLA research has been conducted in the functionalist framework, according to which language consists of bidirectional function-to-form mappings (Cooreman and Kilborn 1991: 197). In this theoretical perspective, linguistic description should account for: a) the grammatical forms, b) the semantic functions (e.g., semantic roles, temporality and aspect etc.) and the pragmatic functions (e.g., topic/focus relationships) encoded by the grammatical forms, and c) the mechanisms governing the grammatical expression of these functions (Giacalone Ramat 2003; Tomlin 1990).

One of the main initial results of functional approaches to SLA has been the definition of a *basic variety*, a simple yet structured stage of interlanguage which “reflects the necessary, rather than the more accidental, properties of the human language capacity” (Klein and Perdue 1997: 304). The basic variety shows a “non-finite utterance organization”, that is, utterances contain verbs and are structured

according to their valency, but there is no trace of inflection. Nouns and verbs occur in an invariant form which corresponds to the stem, the infinitive or the nominative for nouns of the target language. Information about temporality, aspect, person, number, gender may be conveyed by non-inflectional means, such as lexical items (e.g., adverbs, quantifiers), such as the numeral *tre* ‘three’ in (1), which quantifies the uninflected noun *lingua* ‘language’:

(1)	io	parla	tre	lingua
	I	speak:PRS.3SG	three	language:SG
TARGET	<i>io</i>	<i>parlo</i>	<i>tre</i>	<i>lingue</i>
	I	speak:PRS.1SG	three	language:PL
	‘I speak three languages.’			

The emergence of the basic variety represents the overcoming of a pre-grammatical stage characterised by an entirely pragmatic mode (no stable syntax, word order governed by pragmatic principles only, such as topic/focus organisation). Many learners, especially in on-going conditions of low exposure, tend to fossilise at the basic variety stage, which satisfies basic communication needs despite the lack of sophisticated grammatical organisation. Under adequate conditions of exposure, however, the acquisition develops through a series of successive varieties, whose internal organisation is systematic, as is the transition from one variety to the next (see Table 2).

	Variety		
	<i>Pre-basic</i>	<i>Basic</i>	<i>Post-basic</i>
GRAMMATICAL CATEGORIES	None	Verb/Arguments	Verb/Arguments
MORPHOLOGY	None	Basic forms (uninflected)	Inflected nouns and verbs
ORGANISATION OF THE UTTERANCE	Pragmatic	Semantic-syntactic	Syntactic
DEPENDENCY FROM THE CONTEXT	High	<.....>	Low

Table 2. Initial interlanguages (Banfi and Bernini 2003: 84, adapted).

The basic variety approach to SLA has many advantages:

- it allows the analysis of interlanguage per se, as any other natural language and not in relation to the target language;
- despite relevant differences in the basic theoretical assumptions (as well as explicit criticisms, cf. Vainikka and Young-Scholten 2006), it exhibits a high degree of comparability to other approaches relevant to the analysis of LESLLA learners' interlanguages⁵;
- the basic variety approach is one of the main theoretical frameworks of Italian research on SLA, which is in turn the benchmark for data analysis in the research presented here.

It goes beyond the scope of this article to detail the Italian research results. However, an interesting aspect that emerges from many studies is that all learners have difficulties in acquiring morphology, largely due to language internal factors, such as low perceptual salience of (unstressed) morphological endings and (clitic) articles; accumulation of grammatical categories (e.g., gender and number in Ital. *giall-e* 'yellow.F.SG'); homonymy of morphs (e.g., Ital. *-e* 'M.SG / F.SG / F.PL') (Chini and Ferraris 2003; Valentini 2016).

The degree of exposure to the linguistic input, which can be more or less rich and intense, affects the rate of access to L2 grammar and the degree to which grammatical forms are acquired, that is, the outcomes of SLA (with possible persistence of initial stages). Whether or not the lack of alphabetical skills (and, hence, the related ability to explicitly segment the oral speech) leads to a further slowdown or arrest in the acquisition process or produces different results with respect to those described in the existing literature still needs to be verified, first of all by trying to isolate the variables at work in a more explicit way.

⁵ Cf. Vainikka and Young-Scholten who argue that at the beginning of the acquisition path learners do not project any functional syntax, despite the opportunity of transferring this from their L1s. Learners' initial interlanguages are "minimal syntactic trees" whose headedness is based on their L1s. When it comes to acquire functional elements, because they differ across languages, they must be acquired in response to the input. The Organic Grammar's minimal tree has many points of contact – and therefore is compatible for the purposes of the description – with Klein and Perdue's (1997) description of the basic variety as a "non-finite utterance organisation".

3. Corpus design: context, participants and data collection

3.1. The research context

Since 2012, the School of Italian Language for Foreigners of the University of Palermo (ItaStra) has dedicated considerable efforts in organising language and literacy instruction for the local migrant population. Based on agreements with city and national authorities and with various European funding support over the years, migrant learners – a fairly new typology for classes held at universities – have been offered specific literacy paths and/or they have been involved in regular language courses, side by side with the more usual learners, such as international students, PhD and visiting scholars, and professionals. This choice responds to both ethical needs of social inclusion and linguistic needs, as it ensures rich and articulated linguistic input and increases the opportunities for using the target language by learners typically living in contexts of little exposure (Amoruso, D’Agostino and Jaralla 2015).

Data on the linguistic and sociolinguistic composition of the migrants at ItaStra derive from a survey conducted in 2017-2018 for the AMIF project “The strength of the language” (D’Agostino 2018). It involved a sample of 774 migrants, both long-term resident and newly arrived from North Africa, various countries of Sub-Saharan Africa, China and South Asia (especially Bangladesh) and to a lesser extent Europe (Serbia).

Only 15% were monolingual. Plurilingualism especially characterises learners from Sub-Saharan Africa, due to the high degree of societal multilingualism in this area. Individual repertoires may include several languages used with different levels of competence, e.g., languages spoken within the family or used to communicate with neighbouring villages, former colonial languages (still official languages and used in formal education, i.e., French and English), pidgins (e.g., Nigerian Pidgin English) and sometimes language acquired during the migration experience (e.g., Arabic).

While information on plurilingual repertoires was collected through a questionnaire, data on L2 Italian derive from a language test, which revealed a competence from A1 downwards for 58.5% of

the sample. Low competence in Italian also involves 49.22% middle- and long-term residents (D'Agostino and Lo Maglio 2018: 24).

Data on literacy derived from a dedicated test (Amoruso and Lo Maglio 2018). It can be administered in various L1s (Bambara, Mandinka, Pulaar, Wolof) and schooling languages (Arabic, English, French, Italian) and in different writing systems (Arabic, Bangla, Chinese, Hindi, Roman, Tamil). About 31% of the sample was not fully literate in the L1 or in an early learnt language. These data do not overlap with self-declared schooling, as almost 60% of non- or very low-literate individuals declared a short school experience.

Inconsistencies between data on literacy and data on schooling can be explained if we consider the low standards of education in many of the home countries involved, which can hardly guarantee full literacy (UIS/EFA GMR 2015). Moreover, in Western and Middle Africa, learners' native languages are not involved in formal education (thus, in primary literacy), even when these languages have an autonomous writing system (e.g., Wolof). In these areas, education is still entrusted to languages spoken by a small minority of the population, namely English and French (cf. UNESCO 2012). In this context, several years of school attendance are not equivalent to full (and sometimes not even partial) literacy (D'Agostino 2017).

The spread of Qur'anic schools in many countries involved in the survey (especially African countries such as Gambia, Mali and Senegal, as well as Bangladesh) makes the scenario even more complex. In fact, at least in the most traditional schools, education promotes memorisation of the Qur'an in Arabic (that is, a foreign language) through oral repetition. As Saleem (2018: 28) observed, "before starting the actual memorization memorizers are taught Arabic letters and sounds, and how to make 'words' out of them. This learning of 'words', however, is restricted to phonological form in that they do not learn the meaning." In other words, memorisers fluently read a text they do not actually understand. These learning practices, whose purpose is the accurate recitation of the Qur'an, appear to remarkably increase learners' memory capacity and to enhance their prosodic skills (e.g., in reproducing the L2 pitch contour, cf. Maffia and De Meo 2015). However, they can be hardly

considered as literacy practices in the strict sense, nor they promote the acquisition of the Arabic language⁶.

3.2. Selection of the research sample

The research is based on a sample of 20 newcomers recruited during literacy testing at ItaStra in 2017-2018. It involved selection from a larger initial pool of 40 learners, a group which has undergone a significant contraction over time; whenever longitudinal data were not complete, learners were removed from the research sample.

The 20 subjects are young adult learners, male, aged between 18 and 30 years. They arrived from Western Africa (Burkina Faso, Gambia, Guinea, Ivory Coast, Mali, Nigeria, Senegal) and Bangladesh. They speak various African West Atlantic (e.g., Pulaar), Mande (e.g., Bambara, Mandinka), Benue-Congo (e.g., Esan, Igbo, Ika), Gur (e.g., Senufo) languages and Bangla. Especially African learners have plurilingual competence, which may have a functional space in the interaction with other learners.

Plurilingual repertoires still did not include Italian at the time of learners' arrival on the Sicilian coasts, between 10 and 21 months before the first interview. Most of them had attended Italian language courses after arrival, especially in volunteer contexts. This experience, if not too short or discontinuous, was important especially because it provided an opportunity for interaction with the natives, which is in general rare for migrants, and allowed many of them to enter the very initial stages of L2 Italian acquisition. In fact, the presence or absence of this initial linguistic experience marks a gap between those who were (although minimally) able to interact in Italian during the first interview and those who used only languages other than Italian (e.g., English). However, none of them had gone much beyond the non-morphological stage (that is, the basic variety or the immediate post-basic one) despite a length of residence of many months. In fact, apart from possible language courses, their exposure to Italian was on the

⁶ In the sample discussed in this paper (see Section 3.2), learners who attended the Qur'anic schools do not include Arabic among the languages of their repertoires. Learners who report competence of L2 Arabic acquired this language through contact with Arabic-speaking people, frequently during their migratory trip.

whole inconsequential, because of the condition of segregation in hosting centres where migrants live separated from the local community and excluded from any form of true linguistic immersion.

Learners' degree of schooling and literacy in a native or early learnt language varied considerably. Based on the literacy test, learners were ranked on three levels:

- Group 1: no literacy (not able to read and write isolated words in any writing system);
- Group 2: low literacy (recognition of letters/characters, in the Roman alphabet or other writing systems; spelling of words into syllables; linking of multiple syllables; slow deciphering of a few words; writing her/his own name or individual letters/characters or individual words);
- Group 3: moderate to high literacy (fluent reading and writing of simple to complex sentences or texts).

The three levels of literacy did not perfectly correspond to an ideal continuum "no schooling > low schooling (less than 5 years) > schooling (at least 10 years)" (see Minuz 2005: 39). Group 1 also included learners who had only attended a Qur'anic school⁷ or cases where learners had experienced such a low amount of schooling that it left no mark on literacy competence; group 2 included learners who had developed the rudiments of writing and reading (isolated words) outside the school context, typically during their migration or during the initial language courses in Italy; group 3 included learners who had attended high school.

	LITERACY		SCHOOLING
Group 1	No literacy		No schooling 2-5 years of school (any type)
Group 2	Low literacy	early late	10 years of Qur'anic school Experience of literacy in informal and/or volunteer-led contexts
Group 3	Literacy	early	Schooling (8-12 years of high school)

Tab. 3. Literacy and schooling.

⁷ It is interesting to observe that none of these learners included Arabic among the languages of their repertoires. When Arabic is part of the repertoire, as in fact happens in other cases among migrants, it has been learnt through contact with Arabic-speaking people, not infrequently during their migratory journey.

3.3. Data collection

Data were collected longitudinally for 13 months, in 5 individual sessions carried out in non-school settings, i.e., in the hosting centres. Sessions included mainly monological tasks (guided conversations and narrative tasks; see Pallotti et al. 2010).

Session 1 included two steps, that is, a preliminary interview and the ItaStra literacy test. The interview was aimed at collecting sociolinguistic information (e.g., age, origin, linguistic repertoire, schooling and self-reported literacy). Interviews were conducted in Italian whenever possible, otherwise in English; thus, they also provided preliminary information on learners' L2 Italian competence. Data from both the interviews and the literacy test, as well as metadata, were then stored in a digital database.

Session 2 took place the day after Session 1 and was divided into two tasks. First, a short wordless video (adapted from the web) showing in parallel the day of two adults was shown to each learner individually, who was then asked to recount it. The choice of the video depended on the simple, realistic and non-childish character of the story and, above all, on the linear structure of the narrative. It represents a sequence of basic daily activities (e.g., sleeping, cooking, drinking a coffee etc.), which are representative of various event types (action, states, processes) and require a basic vocabulary probably known by learners. The task was aimed at eliciting data on basic lexical categories, that is, whether nouns and verbs: a) were encoded as distinct categories (pragmatic or post-pragmatic phase); b) if distinct, they were represented by basic forms (basic variety) or inflected forms (post-basic varieties). The second task was to select and order a set of images on paper, which once again represented basic daily events, and recount the day before the test with the help of this visual support. The aim was to elicit data on the encoding of temporal-aspectual information (by virtue of the shift to the past) and the verb person (by virtue of the shift to the autobiographical 1st person, instead of the 3rd required by the previous task).

Session 3 took place after about 6 months. Learners were involved in a new narrative-descriptive task. This time, they were asked to talk about their life in Palermo, focusing on an event or a person in particular. This task required more linguistic and communicative

autonomy compared to the previous ones. It aimed at eliciting the same linguistic phenomena as Session 2, namely, nominal and verbal inflection and its possible development after a certain timespan.

Session 4 took place after another 4 months and consisted in two steps. First, there was a conversation about learners' life experience between the third and the fourth sessions, in order to assess their ability to refer to past events and decontextualized experiences and situations. At the same time, the task allowed gathering relevant information about language exposure in the meantime (at school, at work, etc.). The second step was a narrative task based on the wordless video *The pear film* (Chafe 1980). Also in this case, learners were asked to recount the story in order to elicit data on the possible development of their nominal and verbal morphosyntax.

Session 5 was carried out after another 3 months (and 13 months from the beginning of the survey). It included all the tasks already performed in the previous sessions, namely: a semi-guided conversation (on learners' life in Sicily, possible school experience, their native country etc.), the recounting the first video and *The pear film*. Retrieving already used tools was aimed at making data collected at this final stage more straightforward to compare with those elicited in the previous sessions.

The sessions in which data collection was organised are schematically reported in Table 4.

SESSION	ACTIVITIES	DATA ELICITED	TIME	SUPPORT
1	Interview Literacy test	Sociolinguistic information	After 7 to 13 months from the arrival	Paper template
2	Narrative tasks	Presence of nouns and verbs Nominal and verbal inflection	After 7 to 13 months from the arrival	Audio recording
3	Interview Literacy test	Nominal and verbal inflection	After 6 months	Audio recording
4	Interview Narrative task Literacy test	Nominal and verbal inflection	After other 4 months	Audio recording
5	Interview Narrative tasks Literacy test	Nominal and verbal inflection	After other 3 months	Audio recording

Table 4. Synopsis of data collection sessions.

3.4. Data elaboration: transcription and annotation

Sessions 2 to 4 were audio-recorded in .wav format and then manually transcribed using the ELAN software. Each transcription contains as many tiers as there are speakers involved, typically two, i.e., the learner (indicated by an abbreviation) and the interviewer (INT). Transcriptions adopt conversational principles (Jefferson 1984, 2004) in that they provide an orthographic representation of units of language (based on Italian or the other languages used) and also account for speech phenomena, self-corrections and changes in the morphological plan (indicated by the symbol |, as in 2), pauses (indicated by the symbol /:) and interaction (e.g., turn overlap):

- (2) **mio** **amico:** | **amici:** **parla** **wolof**
 my:M.SG friend:M.SG friend:M.PL speak:PRS.3SG Wolof
 ‘I speak Wolof with my friends.’

At this stage, the research does not involve the automatic annotation of morphosyntactic categories, which is a crucial yet still unsolved issue in SLA in general. As Andorno and Rastelli (2009) observed, one of the main problems in adopting a shared system of annotation lies in the inherently unstable character of the interlanguage whose forms are difficult to trace back to expected categories. This makes the creation of tools to perform automatic coding operations extremely difficult. This is especially true if one adopts the internal perspective of interlanguage (working on form-function pairs as they occur in learners’ utterances), rather than measuring its degree of deviation from the target language (cf. Bley-Vroman’s 1983 notion of *comparative fallacy*).

In the absence of a more sophisticated labelling system, which only larger and collaborative research projects can design in the future, interlinear glosses are adopted for data analysis. These are conceived as a space of preliminary reflection on the analytical categories to be used to interpret the interlanguage forms. A mainly internal perspective is adopted, although, for the sake of simplicity, the description of the forms is based on the target language in so far as interlanguage and target language converge at the surface. But it should be observed that morphological glosses only provide a formal

description of the forms produced by learners and are not aimed at interpreting them based on the similarity to hypothetical and reconstructed target forms. This can be observed in (3), in which the lack of number agreement emerges from the labels:

- (3) **mio** **amici:**
 my:M.SG friend:M.PL
 ‘My friends.’

This also involves non-target lexical bases constructed according to target morphology (e.g., *bross-are* < Eng. *brush*, see also Jezek 2005: 187; Rosi 2010: 81).

The basic forms of the verb are described on the basis of the target morphology, since any alternative description would require a preliminary choice on what should be interpreted as a basic form; this is shown in (4):

- (4) **io** **guarda** **film**
 I look:PRS.3SG film
 ‘I watch movies.’

Interlanguage forms which deviate from the target at the phonetic or morphological levels are preserved.

Expected but not encoded forms, e.g., auxiliaries (*noi cucinato pasta* for **noi abbiamo cucinato la pasta* ‘we have cooked pasta’), remain unspecified.

Underspecified labels occur when the interlanguage form lacks morphological information, as in (5), where ‘+’ means ‘interrupted word’ and the zero-marking indicates lack of gender and number information, which is compulsory in Italian (**mia amica*):

- (5) **anche** **lei** **tu+** **amig+**
 also she your:Ø friend:Ø
 ‘She’s your friend too.’

Underspecified labels are also adopted when the morphological category is unclear, as in the case of *comportamenti* in (6), which is formally (i.e., in the target language) a plural noun (‘behaviour-s’) but

The adverb *ieri* ‘yesterday’ is assigned a functional label TEMP, in order to highlight that temporal information, which lacks in the uninflected verbal form, is however lexically encoded in the context.

Functional labels are only tentative at this stage. They have important advantages, but also present some problems. The main problem is that they are not always inherent to specific segments, but seem to have scope over the larger context of the sentence. This seriously limits the possibility of applying this type of analysis to more extended corpora and, more important, of translating it into an automatic annotation system (Andorno and Rastelli 2009: 63). On the other hand, functional labels allow questions to be formulated such as “through which forms is the plural expressed?” (not only “what does the X form express?”, cf. Andorno and Rastelli 2009). In other words, they allow to identify the not (or not fully) grammaticalised encoding of categories which are typically conveyed through morphological means in the target language and are entrusted instead to lexical (or lexical-syntactic means) in the interlanguage. Such linguistic items would go unnoticed if we just observed the form of the words.

This approach seems particularly interesting in the perspective here adopted, as it can bring out possible systematic preferences by LESLLA learners: more specifically, whether LESLLA learners select specific form/function pairs or function words in the input when they start to develop the L2 grammar (Vainikka et al. 2017: 247). For instance, data suggest that both literate and non-literate learners develop interlanguage constructions, that is, non-target analytical tools to express notional categories, which are grammaticalised in the target language, e.g., the ‘be’-construction in (7), but the incidence of such forms seems to be stronger in learners with limited literacy⁸.

4. Discussion and conclusion

The preliminary research activities, i.e., the selection of the sample and the collection of the data, have brought to light some problems that seem to specifically correlate to the LESLLA learner population.

⁸ See Mocciano (2020) for final results of the research, which in fact confirm this preliminary conclusion.

The first problem is methodological and concerns the difficulty of isolating the variables involved in the process of second language acquisition by LESLLA learners, namely L1 literacy and early schooling, on the one hand, and the “LESLLA” condition as a whole and the (quantity and quality of) exposure to the target language, on the other hand. The ItaStra literacy test allowed us to separately measure learners’ degree of literacy and shed light on the complex interrelation of this variable with the self-reported schooling experience. Isolating literacy from the degree of exposure to the target language results in a more problematic task, since all migrants have a very low level of interaction with native speakers. As we have seen in 3.2, the role of literacy appears to be irrelevant at the very beginning of the L2 acquisition path (i.e., in the pre-morphological stages), when the only variable at work is exposure to input. The observation of data suggests that the interrelation of the two variables becomes more complex in successive phases, that is, when morphosyntax starts (or should start) to develop.

A second order of problems that slow down and significantly limit the organisation of the research depends on the extremely high rate of attrition of this learner population. Attrition derives from several causes: first, migrants are subject to frequent relocations or voluntary transfers; second, they are forced into difficult existential situations (exhausting work pace and, at the same time, isolation from the local community) that make them much less accessible than other learner categories. As a consequence, longitudinal research is particularly challenging and requires continuous redesign and a very high degree of flexibility on the part of the researcher (as pointed out by Bigelow and Tarone 2004: 697).

A final problem relates to the process of data elaboration in any learner corpus. Although annotation was not included among the objectives of the research, it was necessary to question the possible theoretical choices for the description of data and, on an experimental basis, it was decided to work manually with a set of functional labels, rather than with morphological labelling only. This option is potentially relevant for the analysis of LESLLA learners’ interlanguage in particular, as it could bring to light possible preferences in selecting specific form/function pairs in the path of

development of grammar, which would remain unnoticed with a pure morphological labelling.

In the near future, it will be necessary to broaden the perspective of the research to include: on the one hand, the design of an automatic annotation tool which takes into account the need to extract functional information to ensure a more fine-grained interlanguage analysis; on the other hand, an increase of the debate, also through concrete collaborations, among researchers working on the acquisition of different L2s by LESLLA learners, in order to reach a more systematic comparison among different acquisition paths eventually providing information on the common features (if any) in LESLLA learners' interlanguage.

References

- Amoruso, Marcello and Angelo Lo Maglio. 2018. I test e la griglia di valutazione. Una sperimentazione continua. In Mari D'Agostino (a cura di), *La forza delle lingue, nella migrazione e nella inclusione* (Strumenti e ricerche 8), 181-183. Palermo: Scuola di Lingua italiana per Stranieri.
- Amoruso, Marcello, Mari D'Agostino and Yousif Latif Jaralla (a cura di). 2015. *Dai barconi all'università. Percorsi di inclusione linguistica per minori stranieri non accompagnati* (Strumenti e ricerche 5). Palermo: Scuola di Lingua italiana per Stranieri.
- Andorno, Cecilia (ed.). 2001. *Banca dati di italiano L2. Progetto di Pavia, CD Rom*. Pavia: Università di Pavia.
- Andorno, Cecilia and Stefano Rastelli. 2009. Un'annotazione orientata alla ricerca acquisizionale. In Cecilia Andorno and Stefano Rastelli (a cura di), *Corpora di italiano L2: tecnologie, metodi, spunti teorici*, 49-70. Perugia: Guerra.
- Atzori, Laura and Stefania Spina. 2009. Corpora di italiano L2: difficoltà di annotazione e trascrizione allargata. In Cecilia Andorno e Stefano Rastelli (a cura di), *Corpora di italiano L2: tecnologie, metodi, spunti teorici*, 93-100. Perugia: Guerra.
- Banfi, Emanuele and Giuliano Bernini. 2003. Il verbo. In Anna Giacalone Ramat (a cura di), *Verso l'italiano. Percorsi e strategie di acquisizione*, 70-115. Roma: Carocci.
- Benazzo, Sandra. 2003. The interaction between the development of verb morphology and the acquisition of temporal adverbs of contrast. In Christine Dimroth and Marianne Starren (eds.), *Information structure and the dynamics of language acquisition*, 187-210. Amsterdam and Philadelphia: John Benjamins.
- Bigelow, Martha and Elaine Tarone. 2004. The role of literacy level in second

- language acquisition: doesn't who we study determine what we know? *TESOL Quarterly* 38(4): 689-700.
- Bley-Vroman, Robert W. 1983. The comparative fallacy in interlanguage studies: the case of systematicity. *Language Learning* 33(4): 1-17.
- Castro-Caldas, Alexandre. 2004. Targeting regions of interest for the study of the illiterate brain. *International Journal of Psychology* 39: 5-17.
- Chafe, Wallace (ed.). 1980. *The pear stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood: Ablex.
- Chini, Marina. 2021. Trent'anni di studio sull'italiano L2. In Margarita Borreguero Zuloaga (ed.), *Acquisizione e didattica dell'italiano: riflessioni teoriche, nuovi apprendenti e uno sguardo al passato*, 51-94. Bern: Peter Lang.
- Chini, Marina and Stefania Ferraris. 2003. Morfologia del nome. In Anna Giacalone Ramat (a cura di), *Verso l'italiano. Percorsi e strategie di acquisizione*, 37-69. Roma: Carocci.
- Cooreman, Ann and Kerry Kilborn. 1991. Functionalist linguistics: discourse structure and language processing in second language acquisition. In Charles A. Ferguson and Thom Heubner (eds.), *Crosscurrents in second language acquisition and linguistic theory*, 194-224. Amsterdam and Philadelphia: John Benjamins.
- Corino, Elisa and Carla Marellò. 2017. *Italiano di Stranieri, Corpora VALICO e VINCA*. Perugia: Guerra.
- D'Agostino, Mari (a cura di). 2018. *La forza delle lingue, nella migrazione e nella inclusione* (Strumenti e ricerche 8). Palermo: Scuola di Lingua italiana per Stranieri.
- D'Agostino, Mari. 2017. L'italiano e l'alfabeto per i nuovi arrivati. *Testi e linguaggi* 11: 141-156.
- D'Agostino, Mari and Angelo Lo Maglio. 2018. Profili dei migranti: competenze linguistiche e alfabetizzazione. In Mari D'Agostino (a cura di), *La forza delle lingue, nella migrazione e nella inclusione* (Strumenti e ricerche 8), 23-29. Palermo: Scuola di Lingua italiana per Stranieri.
- Gallina, Francesca and Barni Monica. 2009. Il corpus LIPS (Lessico dell'italiano parlato da stranieri): problemi di trattamento delle forme e di lemmatizzazione. In Cecilia Andorno e Stefano Rastelli (a cura di), *Corpora di italiano L2: tecnologie, metodi, spunti teorici*, 139-151. Perugia: Guerra.
- Giacalone Ramat, Anna (a cura di). 2003. *Verso l'italiano. Percorsi e strategie di acquisizione*. Roma: Carocci.
- Goswami, Usha and Peter Bryant. 1990. *Phonological skills and learning to read*. Hillsdale, NJ: Erlbaum.
- Huettig, Falk and Ramesh K. Mishra. 2014. How literacy affects the illiterate mind. A critical examination of theories and evidence. *Language and Linguistic Compass* 8(10): 401-427.
- Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In Gene H. Lerner (ed.), *Conversation analysis: studies from the first generation*, 13-31. Amsterdam and Philadelphia: John Benjamins.
- Jefferson, Gail. 1984. Transcript notation. In J. Maxwell Atkinson and John Heritage (eds.), *Structures of social action. Studies in conversation analysis*, ix-

- xvi. Cambridge: Cambridge University Press.
- Klein, Wolfgang and Clive Perdue. 1997. The Basic variety (or: couldn't natural languages be much simpler?), *Second Language Research* 13(4): 301-347.
- Kurvers, Jeanne, Roeland van Hout and Ton Vallen. 2007. Literacy and word boundaries, in Nancy Faux (ed.), *Proceedings of the 2nd annual forum of LESLLA*, 45-65. Richmond, VA: The Literacy Institute.
- Maffia, Marta and Anna De Meo. 2015. Literacy and prosody: the case of low literate Senegalese learners of L2 Italian. In Ineke van de Craats, Jeanne Kurvers and Roeland van Hout (eds.), *Adult literacy, second language and cognition*, 129-148. Nijmegen: Centre for Language Studies.
- Minuz, Fernanda. 2005. *Italiano L2 e alfabetizzazione in età adulta*. Roma: Carocci.
- Mocciaro, Egle. 2021. L'acquisizione delle categorie morfologiche in apprendenti adulti analfabeti. In Margarita Borreguero Zuloaga (a cura di), *Acquisizione e didattica dell'italiano: riflessioni teoriche, nuovi apprendenti e uno sguardo al passato*, 249-268. Bern: Peter Lang.
- Mocciaro, Egle. 2020. *The development of L2 Italian morphosyntax in adult learners with limited literacy*. Palermo: UniPa Press.
- Mocciaro, Egle. 2019. La morfologia nominale nell'italiano L2 di minori stranieri non accompagnati: analfabetismo, acquisizione, didattica. In Bruno Moretti, Aline Kunz, Silvia Natale and Etna Krakenberger (eds.), *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso internazionale di studi della Società di Linguistica italiana (Berna, 6-8 settembre 2018)*, 257-272. Milano: Officinaventuno.
- Onderdelinden, Liesbeth, Ineke van de Craats and Jeanne Kurvers. 2009. Word concept of illiterates and low-literates: worlds apart? In Ineke van de Craats and Jeanne Kurvers (eds.), *Low-Educated Second Language and Literacy Acquisition. Proceedings of the 4th symposium, Antwerp 2008*, 35-48. Utrecht: LOT.
- Palermo, Massimo. 2009. *Percorsi e strategie di apprendimento dell'italiano lingua seconda: sondaggi su ADIL2*. Perugia: Guerra.
- Palermo, Massimo. 2005. L'ADIL2 (Archivio Digitale di Italiano L2). *Studi Italiani di Linguistica Teorica e applicata* 34: 357-373.
- Pallotti, Gabriele, Stefania Ferrari, Elena Nuzzo and Camilla Bettoni. 2010. Una procedura sistematica per osservare la variabilità dell'interlingua. *Studi italiani di linguistica teorica e applicata* 39(2): 215-241.
- Reis, Alexandra and Alexandre Castro-Caldas. 1997. Illiteracy: a cause for biased cognitive development. *Journal of the International Neuropsychological Society* 3: 444-450.
- Reis, Alexandra, Luís Faisca, Susana Mendonça, Martin Ingvar and Karl Magnus Petersson. 2007. Semantic interference on a phonological task in illiterate subjects. *Scandinavian Journal of Psychology* 48(1): 69-74.
- Saleem, Amjad. 2018. Memorization without comprehension: a window onto the 'extremities' of the capability of human brain. *Al-Idah* 36(1): 25-40.
- Sanders, Eric, Ineke van de Craats and Vanja de Lint. 2014. The Dutch LESLLA Corpus. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and

- Stelios Piperidis (eds.), *Proceedings of the 9th international conference on Language resources and evaluation (LREC-2014)*, Reykjavik, May 2014, 2715-2718. European Languages Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/227_Paper.pdf
- Schmidt, Richard W. 1990. The role of consciousness in second language learning. *Applied Linguistics* 11: 129-158.
- Schwartz, Bonnie D. 1993. On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition* 15: 147-163.
- Tarone, Elaine. 2010. Second language acquisition by low-literate learners: an understudied population. *Language Teaching* 43(1): 75-83.
- Tarone, Elaine. 1985. Variability in the interlanguage use: a study on style-shifting in morphology and syntax. *Language learning* 35: 373-403.
- Tarone, Elaine and Martha Bigelow. 2005. Impact of literacy on oral language processing: implications on SLA research. *Annual Review of Applied Linguistics* 25: 77-97.
- Tarone, Elaine, Martha Bigelow and Kit Hansen. 2009. *Literacy and second language oracy*. Oxford: Oxford University Press.
- Tarone, Elaine, Martha Bigelow and Kit Hansen. 2007. The impact of alphabetic print literacy level on oral second language acquisition. In Nancy Faux (ed.), *Proceedings of the second annual forum of LESLLA. Research, policy, and practice*, 99-121. Richmond, VA: The Literacy Institute at Virginia Commonwealth University.
- Tomlin, Russell S. 1990. Functionalism in second language acquisition. *Studies in Second Language Acquisition* 12(2): 155-177.
- Turco, Giuseppina and Miriam Voghera. 2010. From text to lexicon: the annotation of pre-target structures in an Italian learner corpus. In Massimo Moneglia and Alessandro Panunzi (eds.), *Bootstrapping information from corpora in a cross-linguistic perspective*, 1-33. Firenze: Firenze University Press.
- UIS/EFA GMR 2015. A growing number of children and adolescents are out of school as aid fails to meet the mark. Policy paper 22, Fact sheet 31 July 2015. <https://unesdoc.unesco.org/ark:/48223/pf0000233610>
- UNESCO 2012. Why language matters for the millennium development goals. <https://unesdoc.unesco.org/ark:/48223/pf0000215296>
- Vainikka, Anne and Martha Young-Scholten. 2007a. The role of literacy in the development of L2 morpho-syntax from an Organic Grammar perspective. In Nancy Faux (ed.), *Proceedings of the second annual forum of LESLLA. Research, policy, and practice*, 123-148. Richmond, VA: The Literacy Institute at Virginia Commonwealth University.
- Vainikka, Anne and Martha Young-Scholten. 2007b. Minimalism and Organic Grammar. In Simin Karimi, Vida Simiian and Wendy K. Wilkins (eds.), *Phrasal and clausal architecture: syntactic derivation and interpretation*, 319-338. Amsterdam and Philadelphia: John Benjamins.
- Vainikka, Anne and Martha Young-Scholten. 2006. The roots of syntax and how they grow. Organic Grammar, the Basic variety and Processability theory. In Sharon Unsworth, Teresa Parodi, Antonella Sorace and Martha Young-

- Scholten (eds.), *Paths of development in L1 and L2 acquisition*, 77-106. Amsterdam and Philadelphia: John Benjamins.
- Vainikka, Anne and Martha Young-Scholten. 1998. Morphosyntactic triggers in adult SLA. In Marie-Louise Beck (ed.), *Morphology and its interfaces*, 89-113. Amsterdam and Philadelphia: John Benjamins.
- Vainikka, Anne, Martha Young-Scholten, Colleen Ijuin and Samawal Jarad. 2017. Literacy in the development of L2 English morphosyntax. In Marcin Sosinski (ed.), *Alfabetización y aprendizaje de idiomas por adultos: Investigación, política educativa y práctica docente. Literacy education and second language learning for adults: research, policy and practice*, 239-250. Granada: University of Granada.
- van de Craats, Ineke. 2011. A LESLLA corpus: L1 obstacles in the learning of L2 morphosyntax. In Christiane Schöneberger, Ineke van de Craats and Jeanne Kurvers (eds.), *Low-Educated Second Language and Literacy Acquisition. Proceedings of the 6th Symposium, Cologne 2010*, 33-48. Nijmegen: Centre for Language Studies (CLS).
- Valentini, Ada. 2016. La prospettiva psicolinguistica sull'input in L2. In Ada Valentini (a cura di), *L'input per l'acquisizione di L2: strutturazione, percezione, elaborazione*, 25-42. Firenze: Franco Cesati.
- Young-Scholten, Martha and Nancy Strom. 2006. First-time L2 readers: Is there a critical period? In Jeanne Kurvers, Ineke van der Craats and Martha Young-Scholten (eds.), *Low Educated Adult Second Language and Literacy: Proceedings of the 1st Annual Forum*, 45-68. Utrecht: LOT.

