



ORIGINAL ARTICLE

The development of postverbal subjects in L2 Italian: A multifactorial corpus analysis

Andrea Listanti¹  and Jacopo Torregrossa² 

¹University of Cologne, Cologne, Germany and ²University of Frankfurt, Frankfurt am Main, Germany
Corresponding author: Andrea Listanti; Email: andrea.listanti@gmail.com

(Received 7 January 2023; revised 21 December 2023; accepted 6 January 2024; first published online 14 February 2024)

Abstract

Most studies on the acquisition of postverbal subjects (VS) in L2 Italian focus on a limited number of linguistic factors that tend to be associated with the production of VS in L1 (e.g., verb class and subject discourse status). Moreover, they analyze homogeneous groups of learners in terms of proficiency, mostly through controlled experiments. In this paper, we present a cross-sectional corpus study based on a multifactorial analysis of the L2 use of VS structures in semi-spontaneous speech. We analyze the production of VSs by learners of different levels of proficiency (A1–C2), considering linguistic factors that trigger the production of VS in L1, but have been unaccounted for in L2 studies (e.g., agentivity of the subject, syntactic configuration of the sentence, contrastive focus). We use a cumulative link mixed model to show how the features of verbs and subjects in VS structures change across proficiency levels. The results indicate learners' progressive mastery of the mechanisms of assignment of the subject function to the postverbal constituent and increasing sensitivity to contrastive focus as a feature relevant for the use of VS. Furthermore, we observe that psychological verbs associated with the use of VS are produced from the earliest stages of L2 acquisition.

Keywords: adult second language acquisition; language production

Introduction

Several studies on second language (L2) acquisition capitalize on the distinction between different types of interfaces, that is, loci of the interaction between components of the language faculty (Sorace, 2011). On the one hand, internal interfaces involve the integration of information across different linguistic domains (e.g., lexicon–syntax and syntax–semantics). On the other hand, external interfaces concern the interaction between linguistic domains (e.g., syntax) and language-external resources (e.g., discourse constraints and cognitive abilities). The acquisition of Italian verb-subject structures (VS, henceforth) is a privileged viewpoint for testing whether the extent to which L2 speakers master linguistic phenomena involving internal or external interfaces is an indicator of their

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

proficiency level. As we will show in the section “The distribution of postverbal subjects in Italian,” the use of VS depends either on properties at the lexicon–syntax interface (e.g., whether a verb is unaccusative or unergative) or at the interface between syntax and discourse (e.g., whether the subject is informationally new; see Belletti, 2001, 2004). Previous works on L2 acquisition suggest two different acquisitional patterns for VS, depending on the type of interface involved (e.g., Belletti et al., 2007; see the section “Previous studies on the acquisition of VS”). However, most of these studies concern only L2 speakers at advanced levels of proficiency as compared to monolingual controls. Furthermore, they are based on controlled experiments in which only the verb class or the information status of the subject is manipulated. However, there are additional linguistic and discourse factors that may affect the production of VS in Italian by L1 and L2 speakers (e.g., complexity of the noun phrase corresponding to the subject constituent, agentivity, contrast; see the section “The distribution of postverbal subject in Italian”).

The present study aims to contribute to the understanding of L2 acquisition of VS in two ways. First, it follows a cross-sectional approach, examining the use of VS across different proficiency levels, from beginners to advanced speakers. Second, it considers other factors affecting VS production beyond verb class and subject information status. The study is based on semi-spontaneous speech data, which are part of one of the biggest available learner corpora of spoken L2 Italian (see the section “Methodology”). For our analysis, we use a multifactorial annotation schema of L2 speech, which considers verb properties (e.g., verb class) and subject informational features (e.g., givenness) but also includes features associated with the subject constituent that have not been considered in previous studies, such as agentivity and contrast.

We aim to characterize the verb and the subject in VS as clusters of linguistic features and observe which featural configuration is associated with one or the other proficiency level in L2 Italian. In particular, we aim to investigate whether the linguistic features exhibited by VS structures are indicators of learners’ proficiency level. To this aim, we use a cumulative link mixed model with proficiency level as dependent variable and the linguistic features of the verb and subject constituent as predictors. This way, we aim to shed some new light on the developmental trajectory of VS structures in L2 Italian.

The distribution of postverbal subjects in Italian

Italian allows for VS. Rizzi (1982) argues that the availability of VS is related to the *pro*-drop nature of Italian: postverbal subjects are syntactically licensed by a phonetically null element in the specifier of the inflectional phrase (IP). The distribution of postverbal subjects with different verb classes is regulated by both lexical factors and information structure constraints. With unaccusative verbs, VS is the unmarked word order. It is used as an answer to a broad focus question (“What happened?”), as exemplified in (1):

- (1) [Che cosa è successo?]
 What AUX_{PRS,3SG} happened
 “What has happened?”
- È arrivato un treno
 AUX_{PRS,3SG} arrived a train
 “A train has arrived”

The word order in (1) can also be used to mark the subject as new information focus, for example, following a subject *wh*-question such as “What arrived?”.

By contrast, SV(O) is the unmarked word order with unergative and transitive verbs. However, VS tends to be used to mark the subject as new information focus. This is shown in (2a) with the unergative verb *parlare* “speak” and (3a) with the transitive verb *comprare* “buy.” In (3a), the contextually given object *il libro* “the book” is pronominalized in preverbal position. Both (2a) and (3a) are appropriate answers to the preceding subject *wh*-question. Furthermore, it should be noted that Italian allows for the use of SV structures in these contexts, provided that the subject constituent is marked as focus by prosodic means. However, Italian native speakers prefer to mark focus on the subject by using word order (VS) rather than prosody (Belletti et al., 2007; Belletti & Leonini, 2004; Bocci, 2008; Drubig, 2003; Torregrossa, 2012a, 2012b).

- (2) Chi ha parlato?
 Who AUX_{PRS,3SG} spoken
 “Who has spoken?”
- a. Ha parlato Francesco
 AUX_{PRS,3SG} spoken Francesco
 “Francesco has spoken”
- b. #Francesco ha parlato
 Francesco AUX_{PRS,3SG} spoken
 “Francesco has spoken”
- (3) Chi ha comprato il libro?
 Who AUX_{PRS,3SG} bought the book
 “Who has bought the book?”
- a. L’ha comprato Maria
 It-AUX_{PRS,3SG} bought Maria
 “Maria has bought it”
- b. #Maria l’ha comprato
 Maria it-AUX_{PRS,3SG} bought
 “Maria has bought it”

Previous studies have shown that postverbal subjects of unaccusative verbs in broad focus contexts, on the one hand, and of transitive or unergative verbs in new information focus contexts, on the other hand, occupy different syntactic positions. Postverbal subjects of unaccusative verbs are base generated as internal arguments of the verb and, hence, exhibit object-like behavior (Belletti, 1988; Burzio, 1986;

Perlmutter, 1978). Postverbal subjects of transitive or unergative verbs occupy the specifier of a low focus projection dominating the verb phrase (Belletti, 2004).

Several studies based on the analysis of oral and written corpora in L1 Italian show that verb class and information structure are just two of the factors that affect the production of VS (Sornicola, 1994, 1995). For example, Sornicola (1995: 76) shows that VS tends to occur in association with certain types of subordinate clauses such as locative relative clauses or indirect interrogatives, as shown in (4). Notably, this happens also in association with given subject constituents, in spite of the general tendency exhibited by postverbal subjects to express new information focus. For example, the subject constituent *le altezze delle formanti* ‘the pitches of the formants’ in (4) is definite, and hence, most likely given.

- (4) [...] [vedere] come variano le altezze delle formanti
 [...] [see.INF] how vary_{PRS,3PL} the pitches of-the formants
 ‘[...] [to see] how the pitches of the formants vary’

The postverbal subject in (5) is definite, too. In this case, VS seems to be triggered by the occurrence of the locative adverb in sentence-initial position (see the discussion in Sornicola, 1994, 1995).

- (5) Nei menù dei ristoranti alla moda comparvero i piatti
 In-the menus of-the restaurants to-the trend appear_{PST,3PL} the dishes
 forti della cucina texana
 strong of-the cuisine Texan
 ‘The specialties of the Texan cuisine appeared in the trendy restaurants’ menus’

The complexity of the noun phrase corresponding to the subject constituent seems to play a relevant role, too. Complex noun phrases tend to appear postverbally due to their prosodic weight (Quirk et al., 1972; Ross, 1967; Wasow & Arnold, 2003). For instance, the subject constituent in (6) consists of a series of three nouns (*vecchi genitori* ‘old parents,’ *nonni* ‘grandparents’ and *vedovi* ‘widowers’) followed by a relative clause (see Sornicola, 1994: 38).

- (6) Muoiono dimenticati vecchi genitori, nonni, spesso vedovi
 Die_{PRS,3PL} forgotten old parents grandparents often widowers
 che anche in inverno riescono al massimo a
 who also in winter manage_{PRS,3PL} at maximum to
 scambiare due parole con la droghiera
 exchange_{INF} two words with the grocer
 ‘Old parents, grandparents, often widowers, who even in winter at most manage to exchange a couple of words with the grocer, die forgotten’.

Finally, Sornicola (1994, 1995) took into account the semantic features of the verb and the subject of VS. She found that a greater number of VS tends to occur in association with dynamic verbs compared to stative ones. Furthermore, she observed that the subjects of these verbs are mostly inanimate and non-agentive.

Overall, the L1 corpus studies by Sornicola (1994, 1995) suggest the necessity to develop a multifactorial analysis of the use of VS in (semi-)spontaneous discourse,

considering different factors beyond the class of the verb and the information status of the subject.

Previous studies on the acquisition of VS

Several studies have focused on the L2 acquisition of VS as a testing ground for understanding how far parameter resetting takes place in L2 acquisition (Liceras, 1988, 1989; White, 1985). As mentioned in the section “The distribution of postverbal subjects in Italian”, the null subject parameter is associated with a number of properties, including the possibility of omitting overt subjects and allowing for postverbal subjects. Therefore, these two properties are expected to emerge at once, if one assumes that the acquisition of a null subject language by speakers of a non-null subject language involves parameter resetting. However, this hypothesis does not seem to be consistent with the empirical evidence presented in some studies based on grammaticality judgments. For example, Liceras (1989) shows that while L1 English and L1 French speakers accept expletive null subjects in L2 Spanish across all proficiency levels, they tend to prefer VS only in association with unaccusative verbs, particularly at lower levels of proficiency. As for the L2 acquisition of non-null subject languages, White (1985) shows that L1 Spanish/L2 English learners who are able to reject VSs in English do not necessarily comply with the overt subject requirement of the target language, especially at lower levels of proficiency. These findings suggest that the properties associated with the null subject parameter are not necessarily acquired at the same time.

Later research has shown that the study of L2 acquisition of VS should distinguish between different types of VSs, based on the linguistic properties exhibited by these structures in the target languages, including verb class and information structure constraints (see the section “The distribution of postverbal subjects in Italian”). In particular, several studies show that L2 learners of null subject languages like Italian and Spanish may acquire VS in association with unaccusative verbs successfully. In contrast, they exhibit difficulties when using VS to mark the subject as new information focus in association with unergative and transitive verbs. For example, Belletti et al. (2007) compare the production of VS by near-native speakers of Italian with English as L1 and Italian native controls, based on a question–answer elicitation task and a story-retelling task. The results of the study show that in the story-retelling task, the L2 participants are native-like in their production of VS in association with eventive unaccusative verbs. In contrast, in the question–answer elicitation task, they tend to avoid the use of VS to mark the subject constituent as new information focus, independently of whether the verb is unaccusative, unergative, or transitive (see also Belletti & Leonini, 2004 and Dal Pozzo, 2015).

A clear asymmetry between the acquisition of VS with unaccusatives and unergatives or transitives is also reported in Lozano (2006) on L2 Spanish. Based on a grammaticality judgment task, the author notices that VS with unaccusatives, on the one hand, and new information focus subjects, on the other hand, is associated with different acquisition outcomes by L1 Greek and L1 English speakers. Notably, the learners’ L1 does not seem to affect the results, even if Greek is a null subject language in an apparent similar way to Spanish.

Similar evidence emerges also from studies on heritage speakers (e.g., Caloi et al., 2018 on heritage Italian adults; Listanti & Torregrossa, 2023 on heritage Italian children) and attrited speakers (e.g., Tsimpli et al., 2004 on L1 Italian and Greek attrited speakers with English as L2). Moreover, verb class and interface conditions are also shown to affect the timing of acquisition of VS structures among Italian monolingual children, whereby VS emerges earlier in association with unaccusative verbs compared to unergative and transitive verbs (Abbot-Smith & Serratrice, 2015; Belletti & Contemori, 2012; Cairncross & Dal Pozzo, 2022; Lorusso et al., 2005; Vernice & Guasti, 2015).

Overall, the results reported in previous studies reflect the divide between internal and external interfaces introduced in the section “Introduction.” The use of VS with unaccusative verbs in broad focus contexts involves the acquisition of the syntax–lexicon interface. In contrast, the production of VS in association with new information focus subjects involves the syntax–discourse interface, that is, the integration of morphosyntactic and discourse-level information related to the felicitous use of VS. Phenomena related to external interfaces may give rise to difficulties among different types of speakers including advanced L2 learners. These difficulties may lead to the observation that the L2 endstate is non-convergent with the one of L1 speakers and shows optionality even at the near-native level (Sorace, 2005, 2011; Sorace & Serratrice, 2009; Torregrossa et al., 2021; Tsimpli & Sorace, 2006; Rothman & Slabakova, 2018; White, 2009).

Very few studies investigate the acquisition of VS adopting a developmental approach. Two exceptions are the small-scale studies by Bettoni et al. (2009) and Nuzzo (2015) within the Processability Theory framework (Pienemann, 1998, 2005). The authors propose that word orders involving canonical alignment between argument roles (e.g., agent, theme), grammatical functions (e.g., subject, object), and constituent structure (e.g., subject-first and object-second) are acquired earlier than those involving non-canonical alignment. In this sense, SV(O) structures should emerge before (O)VS ones. Furthermore, among the (O)VSs, the authors establish a hierarchy of markedness, with VSs with unaccusatives being the least marked structures, followed by VSs with unergatives and, finally, OVSs with transitives, in which both the subject (postverbal) and the object (preverbal) are in non-canonical position. In spite of the different theoretical assumptions, the observations contained in these studies are consistent with the evidence reported in the studies reviewed previously. Crucially, according to the Processability Theory, the above developmental sequence is not affected by cross-linguistic effects, whereby it should be observed independently of learners’ L1 (Pienemann et al., 2005). The typological distance between learners’ L1 and L2 affects only the speed with which learners proceed from one stage of this developmental sequence to the next. This consideration is particularly relevant for our study, since we were not able to consider learners’ L1 (see the section “The corpus”). In this sense, our study shares with the Processability Theory the effort to identify a developmental sequence in the acquisition of VS holding independently of cross-linguistic effects.

In addition to that, Bettoni et al. (2009) and Nuzzo (2015) make two considerations that have important implications for our study. First, they propose that VSs with psychological verbs of the *piacere*-type should be acquired at the same time as VSs with transitive verbs. Psychological verbs correspond to the verb “like”

in English (Belletti & Rizzi, 1988). In Italian, they are unaccusative verbs selecting an experiencer mapped into an indirect object in sentence-initial position (see *a Gianni* “to Gianni” or *gli* “to him” in (7)) and a theme mapped into the postverbal subject position (see *questo* “this” in (7)).

- (7) A Gianni/gli piace questo
 To Gianni/he_{DAT} like_{PRS.3SG} this
 “Gianni likes this”

In Bettoni et al. (2009)’s terms, these structures involve a non-canonical alignment, which accounts for the lateness of their acquisition. Second, the authors argue that the mastery of subject-verb agreement is a necessary condition for the emergence of non-canonical word orders (see, in particular, Nuzzo, 2015). In other words, agreement between a verb and a postverbal subject shows that learners are able to assign the subject function independently of the pre- vs. postverbal position of the corresponding constituent¹.

To conclude, the studies conducted thus far on the L2 acquisition of VS in null subject languages rely on controlled experiments involving the manipulation of two main factors, that is, verb type and information structure. However, if L2 speakers exhibit difficulties in the integration of information at the sentence and discourse level, this should be mostly visible in the analysis of larger discourse units. Furthermore, previous studies are mostly based on groups of L2 speakers that are relatively homogeneous in terms of their proficiency levels. The few studies involving a developmental perspective (either longitudinal or cross-sectional) have a relatively narrow empirical scope. In this study, we aim to overcome these shortcomings by relying on the analysis of semi-spontaneous speech and adopting a cross-sectional perspective, respectively. We aim to investigate the extent to which the linguistic features associated with VS produced by L2 learners could be used as indicators of their proficiency levels. This way, we want to pin down how L2 learners’ sensitivity to different factors involved in the use of VS in the target language progresses across proficiency levels.

The study

This paper consists of a corpus study whose aim is to analyze the acquisition of VS in L2 Italian in a cross-sectional perspective. We conduct a multifactorial analysis of VSs considering the lexical and semantic properties of the verb and the semantic and information structure properties of the subject. We investigate how far the properties associated with verbs and subjects predict the proficiency of the learner who has produced the corresponding VS structure.

As for the verb, we consider its class (unaccusative, unergative, transitive) and its dynamicity (stative or dynamic). We expect the use of VS with unaccusative verbs to be associated with all proficiency levels, without distinctions. In contrast, the use of VS in association with unergative, transitive, and *piacere*-type verbs is expected to predict higher proficiency levels. In the section “The distribution of postverbal subjects in Italian”, we noticed that in L1 Italian, VS tends to occur in association with dynamic verbs. We investigate whether this tendency is visible in the L2 data,

too. Finally, we consider verb frequency in Italian (see the section “Frequency of the verb”). We expect L2 learners’ use of infrequent verbs with VS to be associated with higher levels of proficiency (Laufer & Nation, 1995; Nation, 2001).

As for the subject, we analyze it in terms of its information structure. We distinguish between discourse-given and focus subjects. While previous studies have mostly considered new information focus, this study also takes into account contrastive focus. This choice is mainly related to the nature of the data that we analyze. Contrastive focus constituents are very likely to appear in semi-spontaneous speech (see, e.g., Baumann & Riester, 2013). In this sense, our study extends the analysis to an information structure category that has not been taken into account in previous studies mainly based on focus marking in answers to *wh*-questions. Based on the literature mentioned in the section “Previous studies on the acquisition of VS”, we assume that the use of VS to mark the subject constituent as focus is particularly hard to acquire in L2 Italian. Therefore, we expect this use of VS to be associated with the highest levels of proficiency.

Furthermore, we analyze speakers’ subject-verb agreement errors in association with VS. This is to examine whether learners master agreement independently of the position of the subject. Agreement between a postverbal subject constituent and the verb is a clear indicator that learners are able to assign the subject function to postverbal constituents (see the section “Previous studies on the acquisition of VS”). We expect L2 learners to acquire this property of Italian progressively. Therefore, we expect the production of fewer subject-verb agreement errors to be a predictor of learners’ increasing proficiency. In relation with this analysis, we will consider how far L2 learners rely on the agentivity of the subject as a cue for subject identification in postverbal position. Several studies have suggested that agentivity plays a crucial role for the assignment of the subject function (cf. Bock & Miller, 1991 and Hale & Keyser, 1993 for a syntactic account of the relationship between agentivity and subject-verb agreement). We expect agentivity to drive the production of VS by L2 learners at lower levels of proficiency. As VS becomes more and more stable at higher levels of proficiency, it should be extended to non-agentive subject constituents. We expect this tendency to be visible in particular with unergative and transitive verbs. In contrast, unaccusative verbs tend to denote changes of states and their subjects are usually non-agentive anyway (Burzio, 1986; Perlmutter, 1978 and Sorace, 2000). Therefore, our analysis will first consider whether the non-agentivity of the subject in VS constructions is a predictor of higher proficiency in L2 Italian. Then, we will show if this holds for all or only for certain verb classes.

Along the same lines, we expect the production of VS with the verb and the subject not adjacent to each other—such as in sentences in which a constituent intervenes between the subject and the verb—to be associated with advanced learners’ production.

Finally, we expect L2 learners at higher proficiency levels to produce a greater number of VSs occurring in subordinate clauses or exhibiting a “complex” subject constituent (see the section “Previous studies on the acquisition of VS”). This may be related to two factors. On the one hand, advanced L2 learners tend to produce more complex structures at both the nominal and clausal level compared to L2 learners at lower levels of proficiency (e.g., Housen & Kuiken, 2009). On the other hand, L2 learners may become more and more sensitive to the complexity of the

Table 1. Overview of the predictions of the study

Type of effect	Hypothesis
Effect of verb class	VS used in association with unaccusative verbs across proficiency levels to the same extent; use of VS with unergative, transitive, and <i>piacere</i> -type verbs as an indicator of increasing proficiency
Effect of verb dynamicity	Use of VSs with dynamic verbs as an indicator of increasing proficiency
Effect of verb frequency	Production of VS with infrequent verbs as an indicator of increasing proficiency
Effect of information structure of the subject	Use of VS to mark the subject as new information or contrastive focus as an indicator of increasing proficiency
Subject-verb agreement errors in VS	Decreasing number of subject-verb agreement errors as an indicator of increasing proficiency
Effect of agentivity of the subject constituent	Emergence of VS with agentive subjects at lower proficiency levels; use of VS with non-agentive subjects as an indicator of increasing proficiency
Effect of syntactic configuration	Increasing number of VSs with constituents intervening between verbs and subjects as an indicator of increasing proficiency
Effect of clause type	Increasing number of VSs in subordinate clauses as an indicator of increasing proficiency
Effect of nominal phrase complexity	Increasing number of VSs displaying complex nominal phrases as an indicator of increasing proficiency

nominal phrase or the type of clause as factors favoring the production of VS in L1 (see the section “The distribution of postverbal subjects in Italian”).

Table 1 provides an overview of the hypotheses of our study.

Methodology

The corpus

The data are drawn from the L.I.P.S. (*Lessico Italiano Parlato da Stranieri*) corpus² (Vedovelli, 2006). The corpus contains orthographic transcriptions of oral texts produced by L2 learners of Italian during the exam for the certification of Italian as a foreign language (CILS). It encompasses 1420 transcripts, corresponding to 100 hours of recorded speech. Each transcript is contained in a different text file. All files are stored in a public folder on the website where the corpus is available. In each file, the following information is reported: date and place of the exam, the identification number of the learners, their proficiency level (from A1 to C2), and number of CILS exams previously taken. Unfortunately, the metadata of the corpus do not include any indication of the L1 of the learners. As a result, an analysis of cross-linguistic effects in the production of L2 Italian is not possible with this instrument. However, this is not problematic for the aims of the present study, which consist in identifying

Table 2. Total number of speakers, transcripts, units, VS occurrences, percentage of VSs on the total number of units, and mean number of VSs produced by each learner for each proficiency level

	A1	A2	B1	B2	C1	C2
Number of speakers	14	18	40	42	40	40
Number of transcripts	19	26	60	61	60	63
Number of units	252	569	1759	2444	2255	2784
Number of VS	26	40	134	155	150	148
%VS	10.3%	7%	7.6%	6.3%	6.6%	5.3%
Number of VSs per learner (mean and standard deviation)	1.86 (1.03)	1.22 (1.17)	3.35 (2.72)	3.69 (2.58)	3.75 (3.22)	3.7 (2.6)

the stages in the acquisition of VS independently from learners' L1 (see the section "Previous studies on the acquisition of VS").

The exams usually consist of two parts, that is, a dialogue between the candidate and the examiner and a monologue by the candidate on a specific topic indicated in the file. Independently of learners' proficiency level, the dialogue takes the form of a roleplay set in an everyday situation, while the monologue involves expressing opinions about different aspects of society. The only difference associated with proficiency level is the duration of the exam (the higher the level, the longer both the dialogue and the monologue). As a result, the texts produced are comparable across proficiency levels, because they were elicited based on the same procedure. The public folder also contains a manual where the criteria for transcription are indicated (e.g., symbols for pauses, unintelligible words, non-verbal communication, and comments by the transcriber; see De Mauro et al., 1993, from which these criteria were adapted).

The data have been collected between 1993 and 2006. For this study, we considered all transcripts collected in 2002, that is, the first year in which the transcripts include all proficiency levels, from A1 and C2 (some transcripts from 2003 have also been analyzed in order to reach a comparable number of speakers—around 40—for each proficiency level from B1 to C2; see Table 2). As for levels A1 and A2, the number of transcripts is considerably lower as compared to the next levels. This is due to the fact that the LIPS corpus was initially designed to contain only texts from level B1 onwards. To be able to analyze as many texts at the A1 and A2 levels as possible, we collected the corresponding transcripts from all the following years until 2006. Nevertheless, the amount of data related to the first two levels of proficiency remains significantly lower than that of the other levels (see Table 2). This is mainly because learners at lower proficiency levels are less productive. Due to this unbalanced dataset, we will consider the data from A1 and A2 separately from the rest of the data (see the section "Production of VS at levels A1 and A2").

For all proficiency levels, we divided the transcriptions into units, based on the occurrence of a finite verb (see Torregrossa et al., 2021). By "units" we mean all the sentences produced by the learners, regardless of the position of the subject (i.e., either SV or VS). Our analysis will consider only the instances of VS. Table 2 reports a description of the dataset, indicating, for each proficiency level, the

number of speakers, the number of available transcripts, the number of units, the number of VSs produced, the percentage of VSs on the total number of units, and the mean and standard deviation of VSs produced per learner. As can be seen from Table 2, the ratio between units and VSs remains relatively constant across proficiency levels. The percentage values related to the production of VSs across proficiency levels seem to indicate that the greatest number of VSs is produced at the A1 level. However, this higher percentage is motivated by the relatively frequent production of *piacere*-type verbs (see the section “Production of VS at levels A1 and A2”) in texts that are relatively short. In fact, the mean number of VSs produced per learner tends to increase across proficiency levels, with the greatest leap being observed between level A2 and level B1.

We collected 653 VSs in total. We did not consider presentative sentences (*c’era un uomo* “there was a man”), due to their formulaic nature and systematic association with new information focus. For the latter reason, we also excluded impersonal passives (*si vendono i libri* “books are sold”)—see Cennamo (1995) and footnote 1. We coded all remaining units for a set of linguistic features related to the verb and the subject constituent, respectively (see the sections “The distribution of postverbal subjects in Italian” and “Previous studies on the acquisition of VS”). In particular, we selected the following 11 features, with the feature “information status of the subject” encompassing three different levels of analysis:

- Verb class;
- Dynamicity of the verb;
- Frequency of the verb in Italian;
- Information status of the subject (newness vs. givenness at both the lexical and referential level and contrastivity);
- Subject-verb agreement errors;
- Agentivity of the subject;
- Syntactic configuration (i.e., occurrence of other constituents in addition to VS);
- Clause type;
- Complexity of the subject constituent

We will explain the criteria for our coding in the following sections. Table S1 in *Supplementary Materials* reports some examples of our coding.

Verb class

We considered five main categories:

- *piacere*-type verbs. We refer to Table S3 of the *Supplementary Materials* for a further classification of this type of verbs aiming to show the extent to which the use of these forms is productive.

- (8) Mi piace il modo di vita degli italiani
 I_{DAT} like_{PRS.3SG} the way of life of-the Italians
 “I like Italians’ way of living”

(SP033B1)³

- Unaccusative verbs

- (9) Quando arriva il compleanno
 When arrive_{PRS,3SG} the birthday
 “When the birthday arrives”

(SP152C2)

- Unergative verbs

- (10) Parlano tutti
 Speak_{PRS,3PL} everybody
 “Everybody speaks”

(SP049B1)

- Transitive verbs

- (11) Me lo ha regalato il mio nonno
 I_{DAT} it_{ACC} AUX_{PRS,3SG} given the my grandfather
 “My grandpa has given it to me”

(SP034B1)

- Copular verbs

- (12) Non è stupida questa donna
 Not be_{PRS,3SG} stupid this woman
 “This woman is not stupid”

(SP068B1)

As mentioned in the section “Previous studies on the acquisition of VS”, *piacere*-type verbs belong to the class of unaccusatives. Nevertheless, we decided to treat them as a separate category. According to the predictions of the Processability Theory, VS with this type of verbs should be acquired later than with other unaccusatives, due to the non-canonical alignment between thematic roles, syntactic functions, and constituent order (see the section “Previous studies on the acquisition of VS”). For the distinction of the other verb classes, we relied on Levin et al. (1995) and Sorace (2000). Auxiliary selection (Sorace, 2000) was adopted as the main criterion to distinguish between unergative and unaccusative verbs. As additional diagnostics to identify unaccusative verbs, we used the *ne*-cliticization test (Belletti & Rizzi, 1981) and the participial absolute test (Loporcaro, 2003). We found only three verbs that could select both “to be” and “to have” as auxiliaries (N: 3; 2 occurrences of *vivere* “to live,” 1 occurrence of *suonare* “to play/ring”). In these cases, the context was examined for disambiguation purposes. For example, the verb *suonare* in the sentence *allora suona la banda locale per rendere più piacevole la riunione* “therefore the local band plays to make the meeting more pleasant” has been classified as unergative because it denotes a controlled process with an animate agentive subject (*la banda* “the band”). This use of the verb is different from the one denoting mere emission as an uncontrolled process that normally takes an

inanimate subject and might select the auxiliary “to be” (e.g., *è suonata la sveglia* “the alarm has rung”; see Sorace, 2000: 863, 874–878). Unaccusative verbs include inherent reflexives (e.g. *Giovanni si arrabbia* “John gets angry,” Cennamo, 1995) and anticausatives (e.g. *Si apre la porta* “the door opens”). Transitive verbs include full reflexives and very few instances of passives (*N*: 11, i.e., 1 at the B1 level, 5 at the C1 level and 5 at the C2 level). We also found six instances of verbs used intransitively (e.g., without a direct object) that nevertheless have a transitive counterpart (*scegliere* “choose,” *pagare* “play,” *cucinare* “cook,” *ballare* “dance,” *battere* “beat,” *insegnare* “to teach”). These verbs were classified as unergatives.

Dynamicity of the verb

We distinguished between dynamic and non-dynamic verbs. Non-dynamic verbs were defined as not involving any physical or metaphorical movement in space. They can refer to inherent characteristics of the subject or unmodifiable states (e.g., *essere intelligente* “to be smart”), temporary conditions (e.g., *avere sete* “to be thirsty”), durative actions (e.g., *pensare* “to think”), or non-durative events (e.g., *accorgersi di* “to realize”; see Bertinetto, 1991). We included this analysis because in L1, the production of VS seems to be sensitive to the semantic properties of the verb (see the section “The distribution of postverbal subjects in Italian”).

- [- dynamic] verb

- (13) A me non diverte questa cosa
 to me not amuse_{PRS.3SG} this thing
 “I am not amused by this thing”

(SP038B1)

- [+ dynamic] verb

- (14) Suona la banda locale
 play_{PRS.3SG} the band local
 “The local band plays”

(SP085B2)

Frequency of the verb

We associated each verb with a measure of frequency in Italian as extracted from the B.A.D.I.P. corpus of spoken Italian (*Banca dati dell'italiano parlato*; Bellini & Schneider, 2003–2019)⁴. In particular, we considered the frequency of occurrence of the corresponding lemma in the corpus.

Information status of the subject (newness vs. givenness at both the lexical and referential level and contrastivity)

We coded each subject constituent for its information status, differentiating between the informational categories “new” and “given” at both the lexical and referential level, on the one hand, and “contrastive” and “non-contrastive”, on the other hand.

We conducted this analysis because both “newness” and “contrast” have been shown to trigger VS in L1 Italian (Belletti, 2001).

For the distinction between “new” and “given,” we relied on the taxonomy proposed in Baumann & Riester (2012, 2013) and Riester & Baumann (2013), which has been tailored specifically for corpus data. The authors distinguish between a referential and a lexical level. The referential level indicates whether the referent corresponding to the subject constituent is mentioned in previous discourse (i.e., given; “ref_given”) or not (i.e., new; “ref_new”). The lexical level indicates whether the expression denoting the subject constituent is mentioned in previous discourse (i.e., given; “lex_given”) or not (i.e., new; “lex_new”). The combination of these possibilities leads to the following four configurations (all examples are adapted from Baumann & Riester, 2013):

- “ref_new” and “lex_new,” if both a referent and its denoting expression are introduced in discourse for the first time, as in (15):

(15) *A colleague* came in.

- “ref_new” and “lex_given,” if a referent is introduced in discourse for the first time by means of an expression used in previous discourse, as “another colleague” in (16):

(16) *A colleague* came in. *Another colleague* went away.

- “ref_given” and “lex_new,” if a referent is mentioned again by means of an expression not used in previous discourse, as “the idiot” in (17):

(17) *A colleague* came in. *The idiot* dropped a vase.

- “ref_given” and “lex_given,” if a referent is mentioned again by means of the same expression, as “the colleague” in (18):

(18) *A colleague* came in. *The colleague* dropped a vase.

At the referential level, we included the label “generic” for generic referents, which can be either given or new at the lexical level. At the lexical level, we included the label “PRO” for pronouns, which, by definition, refer to referents already introduced in discourse.

Additionally, we coded the subjects based on their contrastivity (i.e., “contrastive” vs. “non-contrastive”). “Contrastive” subject constituents were identified based on the following criteria, as defined in Riester & Baumann (2013):

- parallelism with another referent in discourse, as in (19). For example, the constituent *io* “I” in (19) evokes the alternative *un altro* “another” in the next sentence:

- (19) una cosa che penso io, magari per un altro è sbagliata
a thing that think_{PRS.1SG} I perhaps for another be_{PRS.3SG} wrong.F.SG
“something that I think, is perhaps wrong for someone else”.
(SP164C2)

- being in the scope of a focal operator such as *anche* “too” or *solo* “only”, as *anche loro* “they too” in (20):

- (20) E vogliono anche loro.
and want_{PRS.3PL} too they
“And they want it, too”
(SP155C2)

- being the focus part of a cleft construction, as *mio marito* “my husband” in (21):

- (21) È stato mio marito di propormi di lavorare
AUX_{PRS.3SG} been my husband to suggest_{INF-I.DAT} of work_{INF}
“it was my husband who suggested me to work”
(SP166C2)

Subjects which were not coded as “contrastive” were automatically coded as “non-contrastive.”

Subject-verb agreement errors

We coded each VS for the presence vs. absence of a subject-verb agreement error, using the labels “1” and “0,” respectively.

Agentivity of the subject

We distinguished between agentive and non-agentive subjects, as in (22) and (23), respectively (Bambini & Torregrossa, 2010). We performed this analysis in order to understand whether the agentivity feature triggers the assignment of the subject status (see the section “Previous studies on the acquisition of VS”):

- [+ agentive] subject

- (22) Cucineremo tutti e due
cook_{FUT.1PL} both
“Both of us will cook”
(SP099B2)

- [- agentive] subject

- (23) Splende il sole
shine_{PRS.1SG} the sun
“The sun shines”
(SP056B1)

Syntactic configuration

We coded each target sentence based on the word order that it exhibits. For example, VS may be preceded by a constituent or a constituent may occur between the verb and the subject. In particular, we distinguish the following configurations:

- V_S corresponds to a simple VS with no other constituent:

- (24) quando vengono i temporali
when come_{PRS,3PL} the storms
“when the storms come”
(SP114C1)

- CLIT_V_S corresponds to a VS preceded by a direct or indirect-object clitic pronoun or a clitic cluster:

- (25) Che l’ho fatto io
that it-AUX_{PRS,1SG} done I
“that I have made it”
(SP172C2)

- XP_V_S corresponds to a VS preceded by a constituent, for example, a direct or indirect object, an adverb, or a prepositional phrase:

- (26) anche in Russia esistono questi programmi
also in Russia exist_{PRS,3PL} these programmes
“these programmes exist in Russia, too”
(SP156C2)

- V_XP_S corresponds to a VS in which the verb and the subject are separated from each other by an intervening constituent, for example, a direct or indirect object, an adverb, or a prepositional phrase:

- (27) gira un po’ l’economia
turn_{PRS,3SING} a bit the-economy
“economy goes on a little”
(SP166C2)

Clause type

We annotated the type of clause, since VS tends to occur in association with certain types of subordinate clauses (see the section “The distribution of postverbal subjects in Italian” and references therein). We distinguish the following clause types:

- Main clauses:

- (28) Le persone le sceglieva il computer
The people they_{ACC.} choose_{IPFV,3SG} the computer
“The people, the computer chose them”
(SP088C1)

- Complement clauses:

- (29) So che possono partecipare anche loro
 Know_{PRS.1SG} that can_{PRS.3PL} participate.INF also they
 “I know that they can also participate”

(SP091C1)

- Relative clauses:

- (30) Il giocattolo che mi ha regalato mia nonna
 The toy that I_{DAT.SG} AUX_{PRS.3SG} gifted my_{F.SG} grandma
 “The toy that my grandma gave me as a gift”

(SP042B2)

- Adverbial clauses:

- (31) quando sono finite le parate
 when AUX_{PRS.3PL} finished the parades
 “when the parades were over”

(SP041B2)

Complexity of the subject constituent

We coded each subject constituent based on its complexity. We distinguished between four levels of complexity: i) simple noun phrases (i.e., bare nouns or pronouns) or noun phrases preceded by a determiner (DET), with DET including articles, demonstratives, possessives, and quantifiers (see (32)); ii) noun phrases containing a pre- or postnominal modifier (specification, SPEC henceforth), with SPEC including adjectives and prepositional phrases such as *il libro del professore* “the book of the professor,” see also (33); iii) clausal postnominal modifiers (COMP, see (34)); iv) the combination between SPEC and COMP (see (35)). We performed this analysis, since the complexity of the subject constituent has been shown to trigger the use of VS (see the section “The distribution of postverbal subjects in Italian”).

- [DET]_N

- (32) non mi viene la parola
 not I_{DAT} come_{PRS.1SG} the word
 “the word does not come to me”

(SP158C2)

- [DET]_N_SPEC

- (33) che c’hanno i bambini italiani
 that have_{PRS.3PL} the children Italian
 “that Italian children have”

(SP161C2)

- [DET]_N_COMP

- (34) è uscito un bando che permette di
 AUX_{PRS.3SG} gone out a competition notice that allow_{PRS.3SG} to
 fare il servizio civile
 do_{INF} the service civil
 “a competition notice that allows you to do the civil service has been issued”
 (SP169C2)

- [DET]_N_SPEC_COMP

- (35) veniva chiesta una somma di denaro che non possedevo
 ask_{PST.IPFV.3SG} a sum of money that not have_{PST.IPFV.1SG}
 “an amount of money that I didn’t have was requested”
 (SP174C2)

Interrater agreement

The data were coded by the first author of this study. An independent annotator coded 25% of the VJs included in the analysis. Both annotators were Italian native speakers and had previous experience with the analysis of linguistic data. Overall, the inter-annotator agreement percentage is 95.1%. The agreement percentages in the respective categories were the following: 93% for verb class, 92% for verb dynamicity, 90% for information status of the subject at the lexical level, 88.7% for information status of the subject at the referential level, 92.7% for contrastivity, 97.3% for subject-verb agreement errors, 96.7% for agentivity of the subject, 96.7% for syntactic configuration, 98% for clause type, and 100% for complexity of the subject constituent. Whenever the decisions of the two annotators diverged, the annotators discussed their choices until they reached an agreement. If no agreement was reached, the corresponding occurrence was discarded and appeared as “NA” in the final dataset (see Larsson et al., 2020 for a careful consideration of interrater reliability in corpus research).

Analysis and results

The Results section is structured into two main parts. First, we analyze the production of VS by Italian L2 speakers at the levels A1 and A2. Then, we present the results of a cumulative link mixed model related to the production of VS at higher levels, from B1 to C2. As we have already noted in the section “The corpus”, the choice of splitting the analysis into two groups is related to the fact that the number of VJs produced in the first two levels of proficiency is lower than the one produced at the other levels. Considering all data together would result in an unbalanced dataset.

Production of VS at levels A1 and A2

Due to the limited number of units and VS structures available in the first two proficiency levels, we only report descriptive statistics illustrating the main

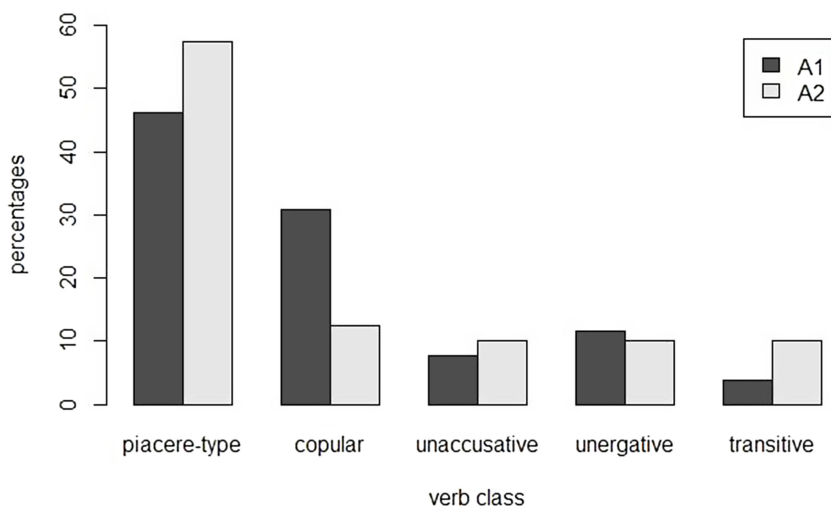


Figure 1. Distribution of VS structures (in percentage) across verb classes (*piacere*-type, copular, unaccusative, unergative, transitive) across the proficiency levels A1 and A2. Percentages are calculated with respect to the total number of VSs produced at each proficiency level.

tendencies which emerged from the data as related to the type of verb. Figure 1 reports, in percentage, the distribution of VSs produced at levels A1 and A2 across verb classes. Most VSs exhibit a *piacere*-type verb (46.2% for A1 and 57.5% for A2, corresponding to 12 and 23 occurrences, respectively). VSs with other verb classes are very infrequent (8 with copulars, 2 with unaccusatives, 3 with unergatives, and 1 with transitives for A1; 5 with copulars, 4 with unaccusatives, 3 with unergatives, and 4 with transitives for A2). Table S3 in *Supplementary Materials* shows that most of the *piacere*-type forms produced by learners at levels A1 and A2 correspond to the use of the first person dative clitic *mi* “to me” and the present form of the third person singular of the verb *piacere*, that is, *piace* (58.33% in A1 and 69.57% in A2).

Production of VS from level B1 to C2

The following analysis includes the data corresponding to the proficiency levels ranging from B1 to C2, that is, 587 VSs annotated for the features described in the section “Methodology”. Our aim is to observe how the featural configuration exhibited by a produced VS is a predictor of L2 learners’ proficiency level. In particular, the statistical model classifies each VS as produced by a learner at one or another proficiency level on the basis of the features exhibited by the verb or subject constituent of the VS itself. Therefore, our outcome variable corresponds to the four proficiency levels considered in this study⁵. Since the corresponding levels are ordered (B1 < B2 < C1 < C2), we used the *ordinal* package (Christensen, 2019) in R (R Core Team 2021) to perform a cumulative link mixed model with proficiency level as dependent variable and the 11 features annotated on VS structures as predictors. Ten of them (i.e., verb class, dynamicity of the verb, information structure at the lexical level, information structure at the referential level, contrastivity, subject-

verb agreement errors, agentivity of the subject, syntactic configuration, clause type, complexity of the subject constituent) are categorical (i.e., either binary or multinomial)⁶. One of them (i.e., verb frequency) is numerical. The values related to verb frequency (see the section “Frequency of the verb”) have been reciprocally transformed according to the Box–Cox transformation (Osborne, 2010).

The main purpose of a cumulative link mixed analysis is to investigate how far the probability for a certain feature to be associated with VS increases (or decreases) across the four proficiency levels (from B1 to C2). Cumulative link mixed models show some advantages compared to ordinal logistic regressions—which also calculate the probability for a certain independent variable to predict an ordinal outcome variable—because they support a random effect structure. In our analysis, we chose the verbal lemma occurring in each VS as random effect⁷. The entire data set and the R script used for the statistical analyses are available at <https://osf.io/mpa2n/>.

Table 3 reports estimates, standard errors (SE), z-values, and *p*-values for each level of the predictor variables. Predictors associated with a significant *p*-value (< .05) are highlighted in bold. In order to distinguish the different predictors, we colored the cells in white or gray.

The results related to verb class reveal that a VS containing a transitive verb (as compared to copular structures in the intercept) is significantly more likely to be produced by a learner with a high level of proficiency rather than a low one. In contrast, we did not find any variation in association with the other verb classes. In other words, while the amount of VSs in transitive constructions tends to increase as learners become more proficient in L2 Italian, we found no evidence that the amount of VSs with unaccusative, unergative, and *piacere*-type verbs varies across proficiency levels. These patterns are shown in Figure 2, which plots the predicted probabilities for the learners of each proficiency level (B1, B2, C1, C2) to produce VS using one of the five verb classes in our analysis (copular, *piacere*-type, unaccusative, unergative, transitive).

As for the dynamicity of the verb, the results do not show any significant variation in the probability for a VS exhibiting a [+dynamic] verb to be classified as produced by a learner at a higher or lower proficiency level. Table 3 also shows a significant effect of verb frequency. VSs are more likely to be classified as produced by more proficient learners if they feature less frequent verbs. The positive estimate is related to the fact that the frequency values had been transformed reciprocally.

The information structure of the subject constituent has been analyzed from two points of view, that is, the “newness” (vs. “givenness”) of the subject constituent—as considered at both the referential and lexical level—and its “contrastivity.” On the one hand, we did not observe any significant variation in the probability for a VS to be classified as produced at a lower or a higher proficiency level based on the lexical givenness (“lex_given”) vs. newness (“lex_new”) of the subject (as compared to a pronoun, in the intercept). Similarly, we did not observe any variation in the probability for a VS to be classified as produced at a lower or higher proficiency level based on the givenness (“ref_given”) or newness (“ref_new”) of the referent corresponding to the subject constituent (as compared to a generic subject, in the intercept). On the other hand, we observed a significant increase in the probability for a VS to be classified as produced by more proficient learners based on the

Table 3. Parameters of the cumulative link mixed model with the learners' proficiency levels as outcome variable and the features associated with VS structures (verb class, verb dynamicity, verb frequency, information status of the subject based on the lexical and referential level, contrastivity, subject-verb agreement errors, agentivity, syntactic configuration, clause type, complexity of the subject constituent) as predictors. The predictors, their estimates, standard errors (SE), and *z*- and *p*-values are given

<i>Predictors</i>	<i>Estimate</i>	<i>SE</i>	<i>z</i>	<i>p</i>
Verb class [<i>piacere</i> -type]	-.12	.6	-.2	.84
Verb class [unaccusative]	.35	.44	.8	.43
Verb class [unergative]	.51	.48	1.05	.29
Verb class [transitive]	1.11	.46	2.39	.02
Dynamicity of the verb [+ dynamic]	.17	.26	.66	.51
Frequency of the verb (reciprocally transformed)	.005	.002	2.56	.01
Information status of the subject at the lexical level [lex_given]	.01	.23	.04	.96
Information status of the subject at the lexical level [lex_new]	-.27	.27	-1.01	.31
Information status of the subject at the referential level [ref_given]	.05	.21	.25	.81
Information status of the subject at the referential level [ref_new]	.04	.23	.19	.85
Contrastivity [contrastive]	.39	.18	2.19	.03
Subject-verb agreement errors [1]	-.74	.33	-2.25	.02
Agentivity of the subject [+ agentive]	-.61	.28	-2.19	.03
Syntactic configuration [CLIT_V_ S]	-.23	.29	-.79	.43
Syntactic configuration [XP_V_ S]	.13	.25	.52	.6
Syntactic configuration [V_XP_S]	1.24	.44	2.84	<.005
Clause type [complement]	.71	.24	2.96	<.005
Clause type [adverbial]	.16	.23	.69	.49
Clause type [relative]	.4	.3	1.34	.18
Complexity of the subject constituent [DET_N_SPEC]	.3	.32	.93	.35
Complexity of the subject constituent [DET_N_COMP]	-.1	.2	-.51	.61
Complexity of the subject constituent [DET_N_SPEC_COMP]	1.24	.5	2.45	.01

contrastivity of the subject. This suggests that the amount of VSs with a contrastive subject increases with proficiency. This is also visible in Figure 3, which plots the predicted probabilities for the learners of each proficiency level (B1, B2, C1, C2) to produce a VS with a contrastive subject.

The results related to subject-verb agreement errors show a significant decrease in the probability for VSs to be classified as produced at high proficiency levels if they feature an error. This suggests that the amount of subject-verb agreement errors produced in association with VS structures decreases as learners' proficiency in L2 increases.

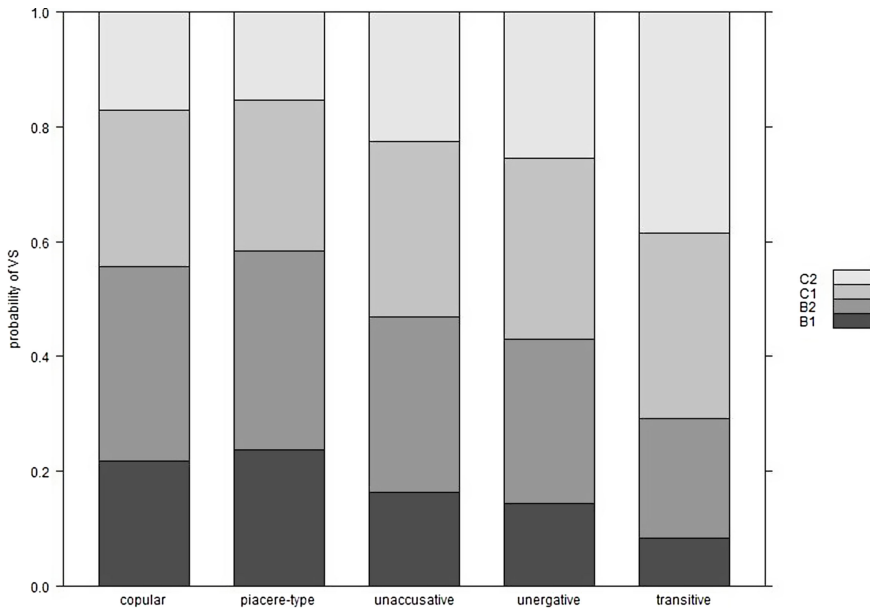


Figure 2. Predicted probabilities for VSs to be classified at a certain proficiency level (from B1 to C2) across verb classes (copular, *piacere*-type, unaccusative, unergative, transitive). The predicted probabilities refer to the model described in footnote 7. The figure has been realized by using the effects package (Fox & Hong, 2009), based on the lattice library (Sarkar, 2008).

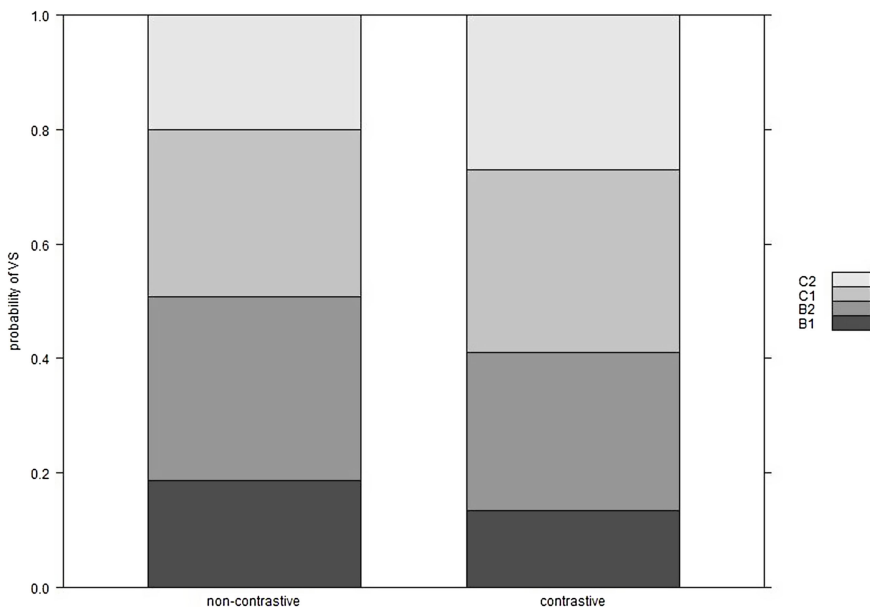


Figure 3. Predicted probabilities for VSs to be classified at a certain proficiency level (from B1 to C2) based on the contrastivity of the subject (0 = non-contrastive; 1 = contrastive). The predicted probabilities refer to the model described in footnote 7. The figure has been realized by using the effects package (Fox & Hong, 2009), based on the lattice library (Sarkar, 2008).

As for the agentivity of the subject constituent, we observe a significant decrease in the probability for VSs to be classified as produced by a learner with a high proficiency level if they feature an [+ agentive] subject. In order to understand whether this pattern varies according to verb class, we considered the number of non-agentive subjects occurring with unaccusatives, unergatives, and transitives, respectively. The percentage of non-agentive subjects with unaccusatives verbs remains stable across proficiency levels. In contrast, the percentage of non-agentive subjects with unergatives and transitives increases from level B2 (33%) to level C1 (38.7%) and peaks at level C2 (71.2%). It should be noticed that the number of unergatives and transitive verbs produced at level B1 (N : 15) is much lower than the one produced at the higher levels (e.g., N : 32 in B2).

The results related to clause type show that the probability for VSs to be classified as produced by a learner with a high proficiency level increases significantly if they occur in a complement clause (as compared to VSs occurring in main clauses in the intercept). This suggests that there is a growing tendency for learners at the higher levels of proficiency to employ the order VS in complement clauses.

As for the complexity of the subject constituent, the probability for VSs to be classified as produced by a learner with a high proficiency level increases significantly if they feature a subject of type “[DET]_N_SPEC_COMP” (as compared to subjects of type “[DET]_N” in the intercept). The same increase is not visible among the other levels of complexity of the noun phrase corresponding to the subject constituent.

Finally, the results related to syntactic configuration show that the probability for VSs to be classified as produced by a learner with a high proficiency level increases significantly if they exhibit the structure “V_XP_S” (as compared to simple “V_S” structures in the intercept). No increase was observed among the other syntactic configurations.

Discussion

Table 4 provides an overview of the results of the study, as related to the predictions formulated in the section “The study” (Table 1).

The first result emerging from our cross-sectional corpus analysis is that there is a significant increase in the probability for a VS to be produced by a learner at a high proficiency level if it features a transitive verb. Notably, the same developmental pattern is not visible in association with unaccusative and unergative verbs, whose use neither increases nor decreases across proficiency levels. In the section “Previous studies on the acquisition of VS”, we showed that VS is the unmarked word order with unaccusatives, but the marked word order with unergatives and transitives. However, the latter two verb types differ from each other in the number of arguments in non-canonical position, that is, one with unergatives (the subject) and two with transitives (the subject and the object). The results of the study suggest that it is not the use of marked structures *per se* that develops with proficiency but rather the use of marked structures involving more than one argument. While the pattern observed with transitives is consistent with the results reported in previous studies, the one observed with unergatives is unexpected when compared with previous

Table 4. Overview of the results of the study as related to the predictions of Table 1 (see the section “The study”)

Type of effect	Hypothesis	Finding
Effect of verb class	VS used in association with unaccusative verbs across proficiency levels to the same extent; use of VS with unergative, transitive, and <i>piacere</i> -type verbs as an indicator of increasing proficiency	No difference in the use of VS with unaccusatives, unergatives and <i>piacere</i> -type verbs across proficiency levels; increased use of VS with transitives in association with higher proficiency
Effect of verb dynamicity	Use of VS with dynamic verbs as an indicator of increasing proficiency	No evidence for a difference in the use of VS featuring dynamic verbs across proficiency levels
Effect of verb frequency	Production of VS with infrequent verbs as an indicator of increasing proficiency	Significant increase in the use of VS with infrequent verbs at higher proficiency levels
Effect of information structure of the subject	Use of VS to mark the subject as new information or contrastive focus as an indicator of increasing proficiency	Significant increase in the use of VS with contrastive subjects at higher proficiency levels
Subject-verb agreement errors in VS	Decreasing number of subject-verb agreement errors as an indicator of increasing proficiency	Significant decrease in the production of subject-verb agreement errors in VS structures with higher proficiency levels
Effect of agentivity of the subject constituent	Emergence of VS with agentive subjects at lower proficiency levels; use of VS with non-agentive subjects as an indicator of increasing proficiency	Significant increase in the use of VS with non-agentive subjects with increasing proficiency
Effect of syntactic configuration	Increasing number of VSs with constituents intervening between verbs and subjects as an indicator of increasing proficiency	Significant increase in the use of VS displaying the syntactic configuration V_XP_S with increasing proficiency
Effect of clause type	Increasing number of VSs in subordinate clauses as an indicator of increasing proficiency	Significant increase in the use of VS in complement clauses with increasing proficiency
Effect of nominal phrase complexity	Increasing number of VSs displaying complex nominal phrases as an indicator of increasing proficiency	Significant increase in the use of VS featuring the most complex nominal phrases with increasing proficiency

research. In the elicited production studies by Belletti et al. (2007) and Caloi et al. (2018), near-native L2 learners and heritage Italian speakers, respectively, behaved like L1 speakers with unaccusative verbs only, while showing difficulties with both unergatives and transitives (see the section “Previous studies on the acquisition of VS”). The findings related to the unergatives observed in the current study may be related to the fact that the class of unergatives occurring in the corpus includes both verbs exhibiting SV and verbs exhibiting VS as unmarked word order (e.g., *telefonare* “to phone,” *chiamare* “to call,” *suonare* “to ring,” and *bussare* “to knock” among the latter; see Benincà et al. 1988). By contrast, previous studies are mainly based on controlled elicited production tasks using only unergatives with SV

as unmarked word order (e.g., *urlare* “to scream” as in Caloi et al., 2018), that is, for which VS is the marked option.

The case of *piacere*-type verbs is worth discussing, too. The Processability Theory predicts that these forms would not emerge at the lowest proficiency levels, due to the non-canonical alignment between thematic roles, grammatical functions, and constituent order (Bettoni et al., 2009; section “Previous studies on the acquisition of VS”). However, our results show that these forms are almost the only occurrences of VS to be found at the levels of proficiency A1 and A2 (Figure 1; cfr. Lorusso, 2014 for similar results). In Table S3 of *Supplementary Materials*, we show that these occurrences almost exclusively exhibit a first person dative clitic (*mi* “to me”) followed by a third person singular verb (*piace* “like-PRS.3SG” in most of the cases, with few occurrences of other verbs of the same class like *serve* “be useful-PRS.3SG” or *basta* “be enough-PRS.3SG”). We speculate that learners at level A1 and A2 use these forms in a formulaic way. The use of these constructions may be favored by the topic of the exam, in which learners are asked to express their opinions. However, it should be noted that a productive use of VS with *piacere*-type verbs cannot be observed at the highest proficiency levels either, except for the use of some inflectional variants of the verb *piace* starting from level B1 (see Table S3 of *Supplementary Materials*). Although we cannot draw any conclusion about the use of VSs with *piacere*-type verbs, this last observation suggests a developmental pattern from a formulaic to a productive use of these verbs. We speculate that the formulaic use of *mi piace* “I like” at levels A1 and A2 may play a pivotal role for the use of VS in later stages of acquisition (Ellis, 2002, 2003, 2012; see also Pallotti, 2007 for a discussion on productivity as a relevant acquisition criterion).

In our analysis of the verbs occurring in VSs, we also considered their semantic properties, distinguishing between dynamic and non-dynamic verbs. Previous studies have not looked at the effect of verb dynamicity on the production of VS in L2. We decided to include this level of analysis based on Sornicola’s (1994, 1995) observation that VS tends to occur in association with dynamic verbs in L1 Italian. We found no evidence of an increase in the probability for a VS to be classified as produced by a learner at higher levels of proficiency if it features a dynamic verb. This suggests that there are certain distributional patterns in the input to which L2ers do not seem to become sensitive (see Ellis, 2002 for a review of studies that share a similar view).

The analysis of verb frequency reveals that VSs featuring infrequent verbs in the L2 input tend to be associated with learners with higher proficiency levels. This result is not surprising given that higher levels of L2 proficiency usually correlate with the development of a richer productive vocabulary (Laufer & Nation, 1995; Nation, 2001). More in general, this result suggests an increasingly productive use of VSs across proficiency levels.

Turning to the analysis of the information structure of the subject in VSs, we observed two tendencies. On the one hand, the use of referentially or lexically new (or given) subjects in VS structures does not seem to affect the likelihood for a certain VS to be classified at a higher or lower proficiency level. On the other hand, VSs featuring contrastive focus subjects tend to be associated with higher proficiency levels. This pattern is unexpected under the hypothesis that L2 learners tend to exhibit difficulties with syntax–discourse interface phenomena across all

proficiency levels (see the section “Previous studies on the acquisition of VS”). However, it appears in line with the concept that the degree of optionality in the production of interface phenomena decreases as the level of L2 proficiency increases. Therefore, our study contributes to the understanding of L2 learners’ mastery of syntax–discourse interface phenomena by showing that proficiency modulates learners’ ability to integrate discourse information in sentence structure (see Sorace, 2011 and references cited in the section “Previous studies on the acquisition of VS”). It should be pointed out once again that the same developmental pattern has not been observed with new information focus subjects. This may be related to the nature of the data, especially because our analysis did not take into account existential constructions, which are used in Italian to introduce new referents in discourse (see the section “Methodology”). At a more speculative level, we do not exclude that different types of focus marking (new information vs. contrastive) may be associated with different developmental paths and outcomes. This result is in line with psycholinguistic studies showing that proficiency modulates the processing of contrastive focus in L2 French. By contrast, a similar effect of proficiency is not observed in the processing of new information focus (Reichle, 2010; Reichle & Birdsong, 2014).

In our analysis, we also considered whether VVs feature subject-verb agreement errors. We found a significant increase in the likelihood for a VS to be classified as produced by a learner with a high proficiency level if it features fewer subject-verb agreement errors. We interpret this result as showing that learners master the mechanisms underlying the assignment of the subject function to postverbal constituents progressively. The results concerning the analysis of the semantic features of the subject in VS are consistent with this conclusion. We found a significant decrease in the likelihood for a VS featuring an agentive subject to be associated with higher proficiency levels. It seems that at lower levels of proficiency, learners consider the agentivity of the subject as a reliable cue for the assignment of the subject function (see the studies mentioned in the section “Methodology”). Once the syntax of postverbal subject constituents is fully in place, the use of VS is extended to less prototypical, non-agentive subjects. This pattern is driven by unergative and transitive verbs, because they allow for both agentive and non-agentive subjects (unlike unaccusatives and *piacere*-type verbs; see the section “The study”).

Additional evidence in favor of a progressive mastery of subject-verb agreement is provided by the analysis of the syntactic configuration in which VS is used. We found a significant increase in the likelihood for a VS exhibiting the structure “V_XP_S”—in which a phrasal constituent intervenes between the verb and the subject—to be classified as produced by learners at higher proficiency levels. This suggests that advanced learners are able to assign the subject function regardless of whether the subject and the verb are adjacent to each other.

With respect to the type of clause in which VS occurs, we observed a significant increase in the likelihood for a VS in a complement clause to be associated with higher proficiency levels. In the section “The distribution of postverbal subjects in Italian”, we noticed that in L1 Italian, VS tends to be produced in association with certain types of subordinate clauses regardless of the information structure of the subject. The tendency observed with L2 learners may be epiphenomenal to a more

general tendency to produce complex syntactic structures at higher proficiency levels. However, it also suggests that L2 learners become more and more sensitive to the association between the use of complement clauses and the production of VS, as observed in the L1. The results related to the complexity of the noun phrase corresponding to the subject constituent could be interpreted along the same lines. We observed that there is a significant increase in the likelihood for a VS to be classified as produced by more advanced learners if it features the most complex nominal phrase type, that is, the one containing both a modifier and a complementizer (see the section “Methodology”). It seems that learners not only produce “complex” noun phrases but also become more and more able to integrate prosodic information—as related to the weight of the subject constituent—in the production of VS (Quirk et al., 1972). However, it is not excluded that once learners start to produce complex noun phrases, they are able to realize them in sentence-final position directly, given that the tendency for prosodically “heavy” constituents to appear in sentence-final position seems to hold cross-linguistically (Arnold et al., 2000; see also Listanti & Torregrossa, to appear for discussion).

In conclusion, the cumulative link mixed analysis carried out in this paper has shed some new light on the development of VS in L2 Italian. Whereas previous studies on L2 acquisition of VS have mainly focused on near-native speakers, our investigation has considered learners of all proficiency levels (from A1 to C2). In particular, we have shown how the linguistic properties associated with the subject and the verb of a VS can be used as an indicator of the proficiency of the learner who produced it. The results related to the verb class and the information structure of the subject are consistent with previous studies: VS with transitive verbs—with both the subject and the object in non-canonical position—and focused subjects seem to emerge later in L2 acquisition. In particular, we looked at the impact of contrastive focus, which has gone unnoticed in previous investigations. Furthermore, our study identifies additional factors affecting the use of VS in L2 acquisition, based on the inventory of linguistic features that have been shown to trigger VS in L1 Italian spontaneous speech. Notably, the inclusion of these features has led us to put the role of verb class and information structure of the subject into perspective. For example, learners seem to be more sensitive to certain features of VS structures, such as the complexity of the noun-phrase corresponding to the subject, the (non-)agentivity of the subject, and the syntactic environment in which they occur, than others, such as the newness of the subject constituent or the dynamicity of the verb (Table 2). In this sense, our analysis underscores the need of a multifactorial analysis of the production of VS in L2 Italian. As mentioned in the section “The distribution of postverbal subjects in Italian,” previous studies on the acquisition of VS in L2 Italian have analyzed the production of VS by L2 learners using controlled experiment, which manipulated one or two conditions at a time (e.g., verb type and information structure of the subject). Our study shows that in L2 spoken production, the effect of these conditions may decrease if other conditions are considered, such as the semantic features of the subject constituent (agentivity) or its syntactic complexity. Likewise, previous literature has identified focus marking as a vulnerable domain in L2 acquisition. Our corpus study allowed us to distinguish different aspects of focus marking (i.e., new information vs. contrastive focus) and identify a different developmental pattern for each of these aspects.

Therefore, our corpus analysis has led us to a more fine-grained description of the interlanguage of L2 learners than the one reported in studies based on controlled experiments. On the other hand, our study would not have been possible without the availability of experimental evidence showing that certain aspects of the L2 acquisition of VS are more problematic than others. In this sense, we strongly believe that a full understanding of the acquisition of certain structures by L2 learners may only be reached triangulating different methodologies, such as corpus analyses with controlled elicitation experiments, with one methodology feeding the other (see Mendikoetxea & Lozano, 2018 for similar considerations).

Limitations of the study

Although the present study has allowed us to advance in the understanding of the acquisition process of VS in L2 Italian, some of its limitations should be taken into account. The most important one concerns the methodology that we have employed. Corpus analyses provide authentic and ecologically valid L2 data but cannot provide negative evidence about the acquisition of certain structures. In other words, we cannot claim that if a structure is not produced (and, hence, does not occur in the corpus), it has not been acquired. As already observed, corpus data need to be triangulated with elicited production and comprehension data, in order to draw more definitive conclusions.

In addition to that, it should be considered that the evidence reported in this paper is based on transcriptions of spoken data. This did not allow us to analyze whether L2 learners relied on additional means to mark information structure, such as prosody. For example, we showed in the section “The distribution of postverbal subjects in Italian” that L1 speakers of Italian may mark a subject constituent as focus by producing SV and associating the subject with a dedicated pitch accent, although focus marking via word order (VS) is the preferred strategy. It is not excluded that the L2 learners considered here relied on prosodic strategies to express information structure distinctions. Therefore, our study allows us to understand the acquisition process of VS in Italian but does not allow us to make a more general claim about L2 learners’ ability to integrate discourse information into sentence structure in sentence production. Moreover, we could not consider the L2 learners’ L1, due to the nature of the data. We do not exclude that cross-linguistic effects from the L1 may modulate the impact of one or the other factor in the production of VS in L2.

Finally, our analysis of the linguistic factors leading to the production of VS in Italian relies on previous studies on L1. However, it should be noted that the data analyzed in these studies consider oral or written texts produced by L1 speakers that are not directly comparable to the ones by the L2 learners considered in this contribution. It is not excluded that text type has an impact on the type of structures selected by the speaker (Pallotti, 2019). Therefore, our conclusion that L2 learners may not be sensitive to certain distributional patterns in the input (e.g., dynamicity of the verb) should be taken with caution. L1 speakers may use a lower number of VSs with dynamic verbs than what is shown in previous studies when producing texts similar to the ones considered in this study. From a methodological point of

view, it would be ideal to compare the data presented here with a comparable native corpus based on the same elicitation procedures (see Lozano, Diaz-Negrillo & Callies, 2020 for the development of a corpus of oral narratives based on both native and L2 data).

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S014271642400002X>

Replication package. The entire data set and the R script used for the statistical analyses are available at <https://osf.io/mpa2n/>.

Notes

1 It should be noticed that Italian allows for structures in which agreement between the verb and the postverbal constituent is optional, that is, *si* impersonal constructions (e.g., *Qui si mangia/mangiano spesso gli spaghetti* “Here the spaghetti are often eaten”; see Cinque 1988, p. 554). These constructions are not included in our analysis (Section “The corpus”).

2 The corpus is freely available online at <https://parlaritaliano.studiumdipsum.it/it/653-corpus-lips>

3 The learners’ codes were assigned by us manually and do not correspond to the more complex identification numbers reported in the corpus (see Table S1 in *Supplementary materials*). In our code, *SP* stands for “speaker”.

4 <http://badip.uni-graz.at>

5 Another way to analyze the corpus data would have been to include SV(O) sentences in the analysis, consider the choice between SV(O) and (O)VS as a binary dependent variable, and include the learners’ proficiency level as predictor in interaction with the 11 linguistic features of our annotation scheme. However, this model would not have been appropriate for our research question, which aims to investigate whether the cluster of features exhibited by each VS is predictive of the proficiency level of the learner who produced it. Our choice to conduct the current analysis is also motivated by practical reasons. Most utterances produced by the learners exemplify the order SV(O). Therefore, the inclusion of SV(O)s in the analysis would have required an arbitrary sampling procedure. It would not have been feasible to apply the same analysis as described in Section “The corpus” to all occurrences of SV(O). Furthermore, the model related to the interaction between proficiency level (with four levels) and the 11 linguistic features (with several levels each) would have been very complex and, hence, difficult to interpret and visualize and could have led to convergence problems.

6 For the predictor variables, the following categories were chosen as the reference level: “copular” for verb class, “[–dynamic]” for dynamicity of verb, “PRO” for information status of the subject at the lexical level, “generic” for information status of the subject at the referential level, “non-contrastive” for contrastivity of the subject, “0” for subject-verb agreement errors, “[– agentive]” for agentivity, “V_S” for syntactic configuration, “main clause” for clause type, “[DET]_N” for complexity of the subject constituent.

7 The resulting model was:

`m <- clmm2(proficiency level ~ verb class + verb dynamicity + verb frequency + information status lexical + information status referential + contrastivity + agreement error + agentivity + clause type + subject complexity + syntactic configuration, random = verb, Hess = TRUE, nAGQ = 10, data = VS_L2)`. For the analysis, we follow the procedure described in Christensen (December 15, 2019), available at https://cran.r-project.org/web/packages/ordinal/vignettes/clmm2_tutorial.pdf. Cumulative link models rely on the proportional odds assumption, according to which the coefficient that describes the relationship between each independent variable and the dependent variable does not change across the levels of the dependent variable (e.g., it is the same for learners at level B2, C1, and C2). However, this assumption is usually difficult to satisfy. Indeed, the Brant test for the full model (Schlegel & Steenbergen, 2020) indicates that the proportional odds assumption is violated ($\chi^2(42) = 98.69, p < .001$). Therefore, in order to make sure that the results were not affected by the violation of the proportional odds assumption, we also conducted a multinomial logistic regression with the same outcome variable (reference level: B1) and predictors as indicated in the text, using *nnet* package (Venables & Ripley, 2002). This analysis relaxes the proportional odds assumption (which is one of the main assumptions of cumulative link models) but does not allow for

random effects. The results of the multinomial logistic regression were consistent with the ones of the analysis described in this paper. The full model corresponding to the multinomial logistic regression is reported in Table S2 of *Supplementary Materials*.

References

- Abbot-Smith, K., & Serratrice, L. (2015). Word order, referential expression, and case cues to the acquisition of transitive sentences in Italian. *Journal of Child Language*, 42(1), 1–31. <https://doi.org/10.1017/S0305000913000421>
- Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55. <https://doi.org/10.1353/lan.2000.0045>
- Bambini, V., & Torregrossa, J. (2010). Cognitive categories behind early Topic/Comment structures. In M. Chini (Ed.), *Topic, struttura dell'informazione e acquisizione linguistica* (pp. 35–58). FrancoAngeli.
- Baumann, S., & Riester, A. (2012). Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning* (pp. 119–162). De Gruyter Mouton.
- Baumann, S., & Riester, A. (2013). Coreference, lexical givenness and prosody in German. *Lingua*, 136, 16–37. <https://doi.org/10.1016/j.lingua.2013.07.012>
- Belletti, A. (1988). The case of unaccusatives. *Linguistic Inquiry*, 19(1), 1–34. <https://www.jstor.org/stable/4178572>
- Belletti, A. (2001). “Inversion” as focalization. In A. Hulk & J.-Y. Pollock (Eds.), *Subject inversion in Romance and the theory of Universal Grammar* (pp. 60–90). Oxford University Press.
- Belletti, A. (2004). Aspects of the low IP area. In L. Rizzi (Ed.), *The structure of CP and IP: The cartography of syntactic structures* (Vol. 2, pp. 16–51). Oxford University Press.
- Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory*, 25(4), 657–689. <https://doi.org/10.1007/s11049-007-9026-9>
- Belletti, A., & Contemori, C. (2012). Subjects in children’s object relatives in Italian. *Revue Roumaine de Linguistique*, 57(2), 117–142.
- Belletti, A., & Leonini, C. (2004). Subject inversion in L2 Italian. *EUROSLA Yearbook*, 4, 95–118.
- Belletti, A., & Rizzi, L. (1981). The syntax of ne: Some theoretical implications. *The linguistic review*, 2, 101–137. <https://doi.org/10.1515/tlir.1981.1.2.117>
- Belletti, A., & Rizzi, L. (1988). Psych-verbs and θ -theory. *Natural Language & Linguistic Theory*, 6, 291–352. <https://www.jstor.org/stable/4047649>
- Benincà, P., Salvi, G., & Frison, L. (1988). L’ordine degli elementi della frase e le costruzioni marcate. In L. Renzi, G. Salvi & A. Cardinaletti (Eds.), *Grande grammatica italiana di consultazione*, Vol. 1 (pp. 129–239). Il Mulino.
- Bertinetto, P. M. (1991). Il verbo. In L. Renzi, G. Salvi & A. Cardinaletti (Eds.), *Grande grammatica italiana di consultazione*, Vol. 2 (pp. 13–161). Il Mulino.
- Bettoni, C., Di Biase, B., & Nuzzo, E. (2009). Postverbal subject in Italian L2: A processability theory approach. In J.-U. Kefler & D. Keatinge (Eds.), *Research in second language acquisition: Empirical evidence across languages* (pp. 153–174). Cambridge Scholars.
- Bocci, G. (2008). On the syntax-prosody interface. *Nanzan Linguistics: Special Issue 5*, 13–42.
- Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23(1), 45–93. [https://doi.org/10.1016/0010-0285\(91\)90003-7](https://doi.org/10.1016/0010-0285(91)90003-7)
- Burzio, L. (1986). *Italian syntax: A government-binding approach*. Springer.
- Cairncross, A., & Dal Pozzo, L. (2022). Postverbal subjects in child Italian: Argument structure, discourse and the definiteness effect. *First Language*, 42(3), 333–360. <https://doi.org/10.1177/01427237211054247>
- Caloi, I., Belletti, A., & Poletto, C. (2018). Multilingual competence influences answering strategies in Italian-German speakers. *Frontiers in Psychology*, 9, 1971. <https://doi.org/10.3389/fpsyg.2018.01971>
- Cennamo, M. (1995). Transitivity and VS order in Italian reflexives. *STUF-Language Typology and Universals*, 48(1–2), 84–105. <https://doi.org/10.1524/stuf.1995.48.12.84>
- Christensen, R. H. B. (2019). ordinal - Regression Models for Ordinal Data. R package version 2019.12-10. <https://CRAN.R-project.org/package=ordinal>.

- Cinque, G. (1988). On Si constructions and the theory of Arb. *Linguistic Inquiry*, 19(4), 521–581. <https://www.jstor.org/stable/4178596>
- Dal Pozzo, L. (2015). *New information subjects in L2 acquisition: Evidence from Italian and Finnish*. Firenze University Press. <https://library.oapen.org/handle/20.500.12657/55229>
- De Mauro, T., Mancini, F., Vedovelli, M., & Voghera, M. (1993). *Lessico di frequenza dell'italiano parlato*. ETASLIBRI.
- Drubig, H. B. (2003). Toward a typology of focus and focus constructions. *Linguistics*, 41, 1–50. <https://doi.org/10.1515/ling.2003.003>
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, N. C. (2003). Constructions, chunking and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Blackwell.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44. <https://doi.org/10.1017/S0267190512000025>
- Fox, J., & Hong, J. (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software*, 32(1), 1–24. <https://doi.org/10.18637/jss.v032.i01>
- Hale, K., & Keyser, S. J. (1993). On argument structure and the lexical expression of syntactic relations. In R. Kayne, R. Zanuttini & T. Leu (Eds.), *An annotated syntax reader: Lasting insights and questions* (pp. 53–109). Wiley-Blackwell.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461–473. <https://doi.org/10.1093/applin/amp048>
- Larsson, T., Paquot, M., & Plonsky, L. (2020). Inter-rater reliability in learner corpus research: Insights from a collaborative study on adverb placement. *International Journal of Learner Corpus Research*, 6(2), 237–251. <https://doi.org/10.1075/ijlcr.20001.lar>
- Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Levin, B., Hovav, M. R., & Keyser, S. J. (1995). *Unaccusativity: At the syntax-lexical semantics interface* (Vol. 26). MIT Press.
- Liceras, J. M. (1988). Syntax and stylistics: more on the “pro-drop” parameter. In J. Pankhurst (Ed.), *Learnability and second languages* (pp. 71–93). De Gruyter Mouton.
- Liceras, J. M. (1989). On some properties of the pro-drop parameter: looking for missing subjects in non-native Spanish. In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 109–133). Cambridge University Press.
- Listanti, A., & Torregrossa, J. (to appear). Matching teaching and learning sequences: L2-acquisition of the interaction between complex noun phrases, word order and information structure. In S. Spina & H. Tyne (Eds.), *Applying corpora in teaching and learning Romance languages*. John Benjamins.
- Listanti, A., & Torregrossa, J. (2023). The production of preverbal and postverbal subjects by Italian heritage children: Timing of acquisition matters. *First Language*, 43(4), 431–460. <https://doi.org/10.1177/01427237231170486>
- Loporcaro, M. (2003). The unaccusative hypothesis and participial absolutes in Italian: Perlmutter’s generalization revisited. *Italian Journal of Linguistics*, 15, 199–264. <https://hdl.handle.net/11384/109528>
- Lorusso, P., Caprin, C., & Guasti, M. T. (2005). Overt subject distribution in early Italian children. In *A supplement to the proceedings of the 29th annual Boston university conference on language development*.
- Lorusso, P. (2014). Verbs in child grammar the acquisition of the primitive elements of the VP at the syntax-semantics interface [Doctoral dissertation, Universitat Autònoma de Barcelona]. <https://ddd.uab.cat/record/127517>
- Lozano, C. (2006). Focus and split-intransitivity: the acquisition of word order alternations in non-native Spanish. *Second Language Research*, 22(2), 145–187. <https://doi.org/10.1191/0267658306sr2640a>
- Lozano, C., Diaz-Negrillo, A., & Callies, M. (2020). Designing and compiling a learner corpus of written and spoken narratives: The Corpus of English as a Foreign Language (COREFL), in C. Bongartz &

- J. Torregrossa (Eds.), *What's in a narrative? Variation in story-telling at the interface between language and literacy*. Peter Lang.
- Mendikoetxea, A., & Lozano, C. (2018). From corpora to experiments: Methodological triangulation in the study of word order at the interfaces in adult late bilinguals (L2 learners). *Journal of Psycholinguistic Research*, 47(4), 871–898. <https://doi.org/10.1007/s10936-018-9560-0>
- Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge University Press. <https://doi.org/10.1017/9781009093873>
- Nuzzo, E. (2015). Ipotesi di sviluppo di ordini sintattici marcati in giovanissimi apprendenti di italiano L2. In M. Chini (Ed.), *Il parlato in (italiano) L2: Aspetti pragmatici e prosodici*. FrancoAngeli.
- Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research and Evaluation*, 15, 12. <https://doi.org/10.7275/qbpc-gk17>
- Pallotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics*, 28(3), 361–382. <https://doi.org/10.1093/applin/amm018>
- Pallotti, G. (2019). Assessing tasks: The case of interactional difficulty. *Applied Linguistics* 40(1): 176–197. <https://doi.org/10.1093/applin/amx020>
- Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. *Proceedings of the Annual meeting of the Berkeley Linguistics Society*, 4, 157–190. <https://doi.org/10.3765/bls.v4i0.2198>
- Pienemann, M. (1998). *Language processing and second language development: Processability theory*. John Benjamins.
- Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of processability theory*. John Benjamins.
- Pienemann, M., Di Biase, B., Hakansson, G., & Kawaguchi, S. (2005). Processability, typological distance and L1 transfer. In M. Pienemann (Ed.), *Cross-linguistic aspects of Processability Theory* (pp. 85–116). John Benjamins.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). A grammar of contemporary English. Longman. <https://doi.org/10.1093/elt/29.1.83>
- R Core Team (2021). R: A language and environment for statistical computing (Version 4.0.5). R Foundation for Statistical Computing [Computer software]. <https://www.R-project.org>
- Reichle, R. V. (2010). Near-nativelike processing of contrastive focus in L2 French. In B. Van Patten & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 321–344). John Benjamins.
- Reichle, R. V., & Birdsong, D. (2014). Processing focus structure in L1 and L2 French: L2 proficiency effects on ERPs. *Studies in Second Language Acquisition*, 36(3), 535–564. <https://doi.org/10.1017/S0272263113000594>
- Riester, A., & Baumann, S. (2013). Focus triggers and focus types from a corpus perspective. *Dialogue & Discourse*, 4(2), 215–248. <https://doi.org/10.5087/dad.2013.210>
- Rizzi, L. (1982). *Issues in Italian syntax*. De Gruyter Mouton.
- Ross, J. R. (1967). Constraints on variables in syntax [Unpublished doctoral dissertation, Massachusetts Institute of Technology]. <https://eric.ed.gov/?id=ED016965>
- Rothman, J., & Slabakova, R. (2018). The generative approach to SLA and its place in modern second language studies. *Studies in Second Language Acquisition*, 40(2), 417–442. <https://doi.org/10.1017/S0272263117000134>
- Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer.
- Schlegel, B., & Steenbergen, M. (2020). brant: Test for Parallel Regression Assumption. R package version 0.3-0. <https://CRAN.R-project.org/package=brant>
- Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, 76(4), 859–890. <https://doi.org/10.2307/417202>
- Sorace, A. (2005). Selective optionality in language development. In L. Cornips & K. P. Corrigan (Eds.), *Syntax and variation: Reconciling the biological and the social* (pp. 55–80). John Benjamins.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1–33. <https://doi.org/10.1075/lab.1.1.01sor>
- Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13(2), 195–210. <https://doi.org/10.1177/1367006909339810>
- Sornicola, R. (1994). On word-order variability: a study from a corpus of Italian. *Lingua e stile*, 29(1), 25–57.
- Sornicola, R. (1995). Theticity, VS order and the interplay of syntax, semantics and pragmatics. *STUF-Language Typology and Universals*, 48(1–2), 72–83. <https://doi.org/10.1524/stuf.1995.48.12.72>

- Torregrossa, J.** (2012a). Towards a taxonomy of focus types: The case of information foci and contrastive foci in Italian. In D. Paperno (Ed.), *UCLA Working Papers in Linguistics, Special Issue on Semantics and Mathematical Linguistics*. University of California.
- Torregrossa, J.** (2012b). Encoding topic, focus and contrast. Informational notions at the interfaces [Unpublished doctoral dissertation, University of Verona].
- Torregrossa, J., Andreou, M., Bongartz, C., & Tsimpli, I. M.** (2021). Bilingual Acquisition of Reference. The role of language experience, executive functions and cross-linguistic effects. *Bilingualism: Language and Cognition*, 24(4), 694–706. <https://doi.org/10.1017/S1366728920000826>
- Tsimpli, I., & Sorace, A.** (2006). Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In D. Bamman, T. Magniskaia, C. Zaller (Eds.), *Proceedings of the 30th annual Boston University Conference on language development* (pp. 653–664). Cascadia Press.
- Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F.** (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8(3), 257–277. <https://doi.org/10.1177/13670069040080030601>
- Vedovelli, M.** (2006). Il LIPS - Lessico di frequenza dell'Italiano Parlato dagli Stranieri. In C. Bardel, & J. Nystedt (Eds.), *Progetto Dizionario Italiano-Svedese. Atti del primo colloquio, Stoccolma, 10-12 febbraio 2005* (pp. 55–78). Romanica Stockholmiensia.
- Venables, W. N., & Ripley, B. D.** (2002). *Modern applied statistics with S* (Fourth edition). Springer.
- Vernice, M., & Guasti, M. T.** (2015). The acquisition of SV order in unaccusatives: manipulating the definiteness of the NP argument. *Journal of Child Language*, 42(1), 210–237. <https://doi.org/10.1017/S0305000913000536>
- Wasow, T., & Arnold, J.** (2003). Post-verbal constituent ordering in English. In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of grammatical variation in English* (pp. 119–154). De Gruyter Mouton. <https://doi.org/10.1515/9783110900019.119>
- White, L.** (1985). The “pro-drop” parameter in adult second language acquisition. *Language Learning*, 35(1), 47–61. <https://doi.org/10.1111/j.1467-1770.1985.tb01014.x>
- White, L.** (2009). Grammatical theory: Interfaces and L2 knowledge. In W. C. Ritchie & T. K. Bathia (Eds.), *The new handbook of second language acquisition* (pp. 49–68). Emerald.