Telic for whom? The Lexical Underspecification Hypothesis

STEFANO RASTELLI [iD]

Department of Humanities, University of Pavia, Pavia, Italy

stefano.rastelli@unipv.it

Abstract

The Aspect Hypothesis (AH) claims that L2 beginners use the perfective morpheme first with telic predicates (e.g., 'arrive', 'build the house') and only later with atelic ones (e.g., 'know', 'work'). In contrast, the Lexical Underspecification Hypothesis (LUH) claims that beginners cannot represent the lexical aspect of L2 predicates (hence the telic vs. atelic distinction), because this distinction is a separate component of verb meaning. To investigate whether L2 learners distinguish between telic and atelic predicates, this study compares the responses from 299 L2 Italian learners (with different L1 backgrounds) and responses from 91 native speakers (NS) to the "*for/in* + time span" adverbial test (Dowty 1979). The analysis shows that native speakers and L2 learners' responses to the adverbial test diverge significantly, with learners' proficiency and – to a lesser extent – L1 modulating their ratings. The results suggest that native speakers and beginning-intermediate L2 learners might not represent telicity alike, either because L2 aspectual competence is still developing or because beginning learners rely on the semantic representations of their L1. These findings support the predictions of the LUH and suggest caution when trying to assess learners' aspectual representations.

Keywords: L2 acquisition, Aspect Hypothesis, telicity, prompted acceptability judgments, L2 Italian

Résumé

L'hypothèse d'aspect (AH) affirme que les débutants en langue seconde (L2) utilisent le morphème perfectif d'abord avec des prédicats téliques (par exemple, « arriver », « construire la maison ») et seulement plus tard avec des prédicats atéliques (par exemple, « savoir », « travailler »). En revanche, l'hypothèse de sous-spécification lexicale (LUH) affirme que les débutants ne peuvent pas représenter l'aspect lexical des prédicats L2 (d'où la distinction télique vs atélique), car cette distinction est une composante distincte de la signification des verbes. Pour déterminer si les apprenants d'une L2 font la distinction entre les prédicats téliques et atéliques, cette étude compare les réponses de 299 apprenants L2 d'italien (avec des langues premières différentes) avec celles de 91 locuteurs natifs au test adverbial « for/in + time span » (Dowty 1979). L'analyse montre que les réponses des locuteurs natifs

et des apprenants L2 au test adverbial divergent considérablement; les compétences des apprenants et, dans une moindre mesure, leur L1 modulant leurs notes. Les résultats suggèrent que les locuteurs natifs et les apprenants débutants-intermédiaires en L2 ne se représentent pas la télicité de la même manière, soit parce que la compétence aspectuelle en L2 est encore en développement, soit parce que les apprenants débutants s'appuient sur les représentations sémantiques de leur L1. Ces résultats confirment les prédictions de la LUH et invitent à la prudence lorsque l'on tente d'évaluer les représentations aspectuelles des apprenants.

Mots clés: acquisition L2, Hypothèse de l'aspect, télicité, jugements d'acceptabilité assistés, Italien

1. INTRODUCTION

The current study investigates whether Italian L2 Italian learners have native-like representations of the lexical aspect of predicates they already use and comprehend, and, specifically, of the telic vs. atelic distinction.[1] In the last thirty years, many authors have used what are called aspectual diagnostics in order to code the lexical aspect of L2 predicates. In particular, in many studies, the lexical aspect of the predicates uttered by learners in production tasks – or presented to learners in comprehension tasks – was determined by means of the adverbial test. In production tasks, if the predicates used by the learners were deemed acceptable with '*in/for* x-time span' expressions, they were coded as 'telic' or 'atelic' respectively. This implies that L2 learners and native speakers (NS) represent the lexical aspect of predicates alike. The current study questions this idea and compares native speaker and learner ratings of supposedly acceptable and unacceptable telic and atelic combinations with '*in/for* x-time' adverbials. Ratings of such combinations may converge or diverge. If they converge, then it is possible that even beginning L2 learners can discriminate between telic and atelic predicates. If they diverge, then it is possible that beginning L2 learners ignore or disregard the fact that L2 predicates are telic or atelic. The idea that L2 learners' aspectual representations are largely incomplete and that beginners may ignore or disregard whether predicates are telic or atelic contradicts the Aspect Hypothesis (AH) and the Lexical Underspecification Hypothesis (LUH).

2. THE ASPECT HYPOTHESIS (AH)

The acquisition of L2 Aspect is one the most investigated issues in the history of Second Language Acquisition (SLA) research (for a recent review of studies, see Bardovi-Harlig and Comajoan-Colomé 2020). One well-known hypothesis that proposes to explain the acquisition of L2 Aspect is known as the 'Aspect Hypothesis' (AH). The AH proposes that L2 learners acquire tense and aspect morphology of

---

[1]Abbreviations used: AH: Aspect Hypothesis; CLIP: Corpora e Lessico di Italiano Parlato; L1: first language; L2: second language; LIP: Lessico dell'Italiano Parlato: LUH: Lexical Underspecification Hypothesis; NS: native speaker; NNS: non native speaker; SLA: Second Language Acquisition; TL: target language.

the target language (TL) asymmetrically, with some verbs[2] learned earlier than others (Andersen and Shirai 1994, 1996; Salaberry and Comajoan 2013). In this approach, the acquisition of the perfective–imperfective distinction is modulated by learners' knowledge of lexical aspect of TL verbs. Lexical aspect is an inherent semantic property reflected in a speaker's representation of the internal temporality of verbs. The representation is inherent because it holds regardless of the morphological representation of events as complete or incomplete, in languages that encode such a distinction (Filip 2012: 721). The best-known classification of lexical aspect, discussed by Comrie (1976) and included in Dowty's (1979) influential semantic decomposition model, follows the Vendler-Mourelatos classification (Vendler 1967, Mourelatos 1978). This divides verbs into four categories – states, activities, accomplishments, and achievements – depending on the clustering of the semantic features [±static], [±durative], and [±telic]. This feature specification can expand further depending on whether the verb is also [±punctual], [±semelfactive], [±ingressive], [±iterative], [±inchoative], etc. (Dik 1994: 29, Tatevosov 2002: 320). The current study focuses only on the [±telic] opposition. Following Dowty (1986: 42-43), the label 'telic' in this study includes both accomplishments and achievements.

The AH claims that learners associate the perfective morpheme with telic verbs first, and only later with atelic verbs. As Andersen (2002: 78) puts it, "learners first use past marking (e.g., English) or perfective markings (Chinese, Spanish, etc.) on achievements and accomplishment verbs, eventually extending its use to activity and then to stative verbs". The opposite pattern holds for the imperfective morpheme. While lexical aspect constrains the early emergence of perfective and imperfective verb morphology, at later stages of L2 acquisition, verb morphology spreads regardless. However, the spread of imperfective morphology from atelic to telic verbs seems not to be equally well attested (Bardovi-Harlig 2000). As Andersen (2002: 78) explains, the AH does not mean that "just any token of an achievement verb will automatically attract a perfective marker, but rather that when the perfective marker is used, it will be used more often with achievements (or perhaps with telic events in general)". Many authors observe that assigning a verb to any of the aspectual classes is a function of the arguments of the verb, thus agreeing with Verkuyl (1993) that lexical aspect is compositional. This means that membership in a lexical aspect category is actually a property of the predicate and that it is co-determined by different factors: external (agency/animacy of the subject) and internal (e.g., cardinality of the object arguments and of the adverbial or prepositional adjuncts). Others highlight the importance of discourse-pragmatic factors and narrative backgrounding or foregrounding (Bardovi-Harlig 1995). It is unlikely that an L2 beginner's competence encompasses all these layers of contextual information, and

---

[2]In this article, the term 'verb' refers to both the lemma and all the lexemes of that lemma, whether they are in the present or past tenses, perfective or imperfective. In contrast, the term 'predicate' indicates the compositional nature of the verb phrase, the fact that the belonging of a verb to a given aspectual category is often – not always – a function of its syntactic environment (number and type of external and internal arguments and adjuncts). The term 'verb' referring to lemmas and lexemes is also found in most earlier formulations of the AH.

learners may instead rely on some core aspectual properties of verbs that they carry over in any communicative contexts, maybe from their first language (Salaberry and Martins 2014: 340).

Different positions exist within the AH framework. For instance, early perfective morphemes do not indicate aspectual categories but are used to oppose present and past tenses (Salaberry 2008); verb dynamicity is more important than telicity (Dominguez et al. 2013); increased use of prototypical pairings goes hand in hand with increased L2 proficiency (McManus 2013). Over the years, two different components became recognizable in the AH: the descriptive and the explanatory components. With respect to the descriptive component, most production and comprehension data from different languages show that the lexical aspect of verbs might influence the acquisition of the L2 Tense-Aspect system. As for the explanatory component, the AH seeks the principles that lead L2 learners to rely on lexical aspect when acquiring the L2 Tense-Aspect morphology. Most scholars agree on the former component (the description) but disagree on the latter (the explanation). Disagreement concerns how L2 learners – especially beginners – can distinguish between telic and atelic verbs of the target language. Where does this knowledge come from? If learners attach the perfective morpheme first to telic L2 verbs, they must somehow represent such verbs as telic. Representing telicity means that L2 learners, similar to native speakers, can represent some verbs as having a virtual resulting state, or 'culmination point'. When thinking of events described by telic predicates, such as 'arrive', 'paint the wall' or 'build the house', even beginning L2 learners should be capable of imagining a culmination point (e.g., something or someone reaching a destination, the wall being completely painted, or the house being completely built). In contrast, atelic predicates, such as 'push a cart' or 'take a stroll', lack a virtual culmination point (Dowty 1979; Krifka 1992, 1998).

A possible source of learners' knowledge of the telic vs. atelic contrast is the distributional bias in the target input. In the emergentist approach to the AH (Li 2000, 2002), it is stressed that the early associations between aspectual morphemes and lexical categories are above all a consequence of learners' implicit capacity to analyze and record the probability of the co-occurrence of forms and meanings in the input they are exposed to. As noted, perfective morphemes are most often associated with telic verbs in the input. The frequency of these co-occurrences activates a number of dynamic, adaptive-associative patterns (gradually more and more generalizable), which in turn provide the ground for shaping the semantic categories underlying the four Vendlerian classes (Li 2000: 309). The semantic categories of lexical aspect would emerge as a result of a bottom-up, non–rule-driven process of acquisition (Li 2002: 84). L2 learners would acquire the telic/atelic distinction because frequent associations reinforce a critical number of neural networks. These neural networks instantiate the Vendlerian categories in a speaker's and a learner's mental grammar. The prototypical associations are therefore both determined top-down and reinforced bottom-up – that is, statistically. Wulff et al. (2009) compare telicity ratings and the lexeme–morpheme association scores extracted from learner corpora. Using multiple distinctive collexeme analysis and unidirectional contingency-based measure 'delta-pi', they find that the verbs that emerge and are

learned first in the progressive aspect are highly atelic, frequent in and associated with the progressive in the input. Likewise, the verbs first learned in the past tense are highly telic, frequent in and distinctly associated with the past tense in the input (Wulff et al. 2009: 104).

Another possible source of learners' knowledge of the telic vs. atelic contrast is the cognitive 'prototypicality principle' (Andersen and Shirai 1994: 133, Andersen and Shirai 1996: 532, Andersen 2002: 81). Prototypical form-meaning associations are perceived as more natural and congruent, so learners acquire them first. The association between telicity and perfectivity is prototypical because terminativeness and boundedness are congruent concepts. The inherent culmination point that characterizes telic verbs (such as the Italian verb *cadere* 'fall') is more acceptable with bounded events presented at the past perfective (e.g., the Italian "passato prossimo" *è caduta* '(she) has fallen/(she) fell') than with unbounded events presented at the past imperfective (e.g., the Italian *imperfetto 'cadeva'* '(she) fell/was falling').

Finally, another source of learners' capacity to distinguish telic from atelic verbs may be semantic (aspectual) transfer from their L1. Beginning L2 learners may project the aspectual values of L1 verbs onto L2 verbs. These values may or may not coincide, depending on the L1–L2 pairing. For example, L1 Spanish–L2 Italian learners could assume that *arrivare* 'arrive' and *cadere* 'fall' are telic because the corresponding verbs in Spanish (*llegar* and *caer*) are telic. However, verbs with very similar meaning may belong to different aspectual categories. For example, Nishi (2012: 398) reports that while in English 'know' is stative, in Japanese the corresponding verb 'siru' is an achievement (meaning 'come to know'). In contrast, the English 'fall' and the corresponding Japanese 'otiru' are both telic. Many studies on L1 transfer effects in the acquisition of the tense-aspect system concern L1–L2 similarities and differences in the domain of grammatical aspect. For example, the list provided by Bardovi-Harlig and Comajoan (2020: 15–20) is replete with studies focusing on the influence of a [+ progressive] L1 vs. a [-progressive] L1 (e.g., English vs. German or Russian) on the acquisition of progressive statives, or on the influence of the French *passé composé* vs. the English present perfect on the acquisition of achievements and accomplishments in various L2.

Other studies focus specifically on semantic transfer, that is, the effect of positive vs negative transfer of lexical aspect categories. Shirai (2013) focuses on the issue of crosslinguistic comparison and suggests that discrepancies between lexical categories influence learners' form-meaning mappings. Shirai (2013) concludes that the predictions of the AH should be modulated by L1 influence, namely, L2 learners would first acquire form-meaning mapping of L2 verbs whose lexical aspect is identical to the corresponding L1 verbs. Nishi (2016: 37) acknowledges that "it is not clear exactly what learners know about the lexical aspect of a particular verb in L2 at a given point" and examines whether and how L1 verb semantics influence the acquisition of aspectual distinctions (progressive vs resultative) of -teiru constructions in L2 Japanese. The starting hypothesis is that L2 learners might acquire tense-aspect markers in a verb-specific manner (based on L1–L2 aspectual similarities) rather than in a rule-based manner (based on the general predictions of the AH). In her studies, the author finds that even advanced learners of Japanese find it difficult to

correctly reject items that are not possible in L2 when there is a cross-linguistic discrepancy in lexical aspect. Nishi and Shirai (2019) utilize oral picture descriptions and find that L2 Japanese learners (with either Chinese, English or Korean as their L1) have difficulty in rejecting incorrect L2 aspectual pairings (but not in accepting correct ones) when the lexical aspect of the L1–L2 translational equivalents of a given verb belong to different aspectual categories.

3.   THE LEXICAL UNDERSPECIFICATION HYPOTHESIS (LUH)

The AH takes for granted that beginning learners can represent and project the lexical aspect values of the verbs they use and comprehend, whether such values are deduced from the distributional properties of verbs or transferred from their L1. There is another logical possibility: L2 learners at early stages of acquisition may not be able to distinguish between telic and atelic verbs. This claim was put forth in the Lexical Underspecification Hypothesis (LUH) (Rastelli 2008, 2009, 2019, 2020a, 2020b, 2021; Rastelli and Vernice 2013). The LUH states that there could be a developmental pattern which constrains the ways learners can represent the lexical aspect of TL verbs over time. Unlike NSs, L2 learners may initially ignore or disregard telicity because the acquisition of the lexical aspect, as well as the whole L2 tense-aspect system, is still developing. Similar to mature native speakers, adult L2 learners can certainly represent the lexical aspect of any events designated by L2 verbs, but they may find it difficult to map lexical aspect values onto newly acquired verbs in online language comprehension and production. If in language processing the category of lexical aspect is temporarily disabled, learners would simply use verbs for their general meaning. In recent years, this idea started to spread even among the proponents of the AH. For example, the 'Lexical Insensitivity Hypothesis' claims that beginning L2 learners are insensitive to lexical aspect. As proficiency improves, learners become more sensitive and produce tense-aspect markers based on the verbs actional templates (Tong and Shirai 2016). However, if aspectual competence only emerges late, beginning L2 learners should not be able to distinguish whether a L2 verb is telic or not.

Evidence for the LUH came from qualitative and statistical analysis of learner corpora, elicited narratives and online processing studies. Qualitative analysis, both longitudinal and cross-sectional, shows that learners regularly overextend the few verbs they already know and already use (the so-called 'basic' verbs, in Viberg 2002) regardless of these verbs' telicity. In sentence (1) (taken from a learner corpus of written story retellings, an L1 English–L2 Italian beginner systematically uses the basic, frequent verb of motion *andar*e 'go' instead of *venire* 'come', although *venire* is telic and deictic, while *andare* is not:

(1)   *Quando io vado qui   io posso vedere miei amici
      When I 1SG-go here I can   see   my friends
      'When I come here I can see my friends.'

In the same corpus of L2 Italian written productions by L1 English learners, beginner and low-intermediate learners often use *andare* instead of less frequent, more

aspectually specified motion verbs such as *salire* 'get on', *arrivare* 'arrive', *avvicinarsi* 'get closer', *allontanarsi* 'move away', or *raggiungere* 'reach'. Verbs which are similar in meaning but different in lexical aspect can also be expected to be interchangeably used by learners at this stage. For example, beginner and low-intermediate L1 Chinese–L2 Italian learners, observed during an oral retelling task, recurrently swap atelic *guardare* 'look / watch' with *vedere* 'see', which – in its basic meaning of 'perceiving that something entered the visual field' – is telic. These learners also often interchangeably use *sapere* 'know (that)' and *conoscere* 'know', as well as *dire* 'say' and *parlare* 'talk'. Finally, learners often switch verbs constituting a 'phasal pair', that is, verbs that represent different phases of the same event, regardless of frequency, designated as 'reversive verbs' according to Cruse (1997). For example, beginner and low-intermediate L1 Chinese, L2 Italian learners misuse *cercare* 'search' and *trovare* 'find'; *insegnare* 'teach' and *imparare* 'learn'; and *dare* 'give' and *ricevere* 'receive'. Although somewhat expected, given the poverty of vocabulary, these observations suggest three things. First, early verbs might be lexically underspecified: in early interlanguages, unlike in mature languages, verb meaning can be temporarily dissociated from its lexical aspect. Second, SLA researchers should not take the aspectual content of those verbs for granted and should avoid classifying verbs at their face value. Third, production and corpora-based data alone are often unreliable for tapping into learners' aspectual competence, because they are sensitive to contextual factors.

The distributional explanation of the AH claims that most early L2 perfectives are telic, highly contingent (i.e., the lexeme and the perfective morpheme are strongly associated), and frequent. Rastelli (2020a) utilizes the contingency-based, unidirectional association score delta-pi to track the emergence of the perfective morpheme in the Corpus Pavia, the largest and best-known longitudinal learner Italian corpus to date (~700,000 tokens, ~15,000 types overall (Giacalone Ramat 2003).[3] The study aims to find out whether early L2 perfectives are contingent upon telicity of verbs, and whether distribution of perfectives in the Italian input[4] affects the patterns of morpheme emergence. Results show that (a) the acquisition of the perfective is not contingent upon telicity but is affected by actional underspecification, generality of meaning, and contextual relevance of verbs; (b) the distribution of perfectives in

---

[3]The corpus Pavia was collected from the mid-1980s to the late 1990s in Northern Italy. It contains transcriptions of about 120 hours of oral interviews of 22 Italian L2 learners from 11 different L1 backgrounds from five typological families; learners (aged 12–48 years) had different length of instruction and residence. Proficiency spanned from beginner to high-intermediate. Learners engaged in spontaneous and semi-structured conversations and tasks with Italian interviewers on a wide variety of topics.

[4]The distribution of perfectives in the Italian input was calculated from the normalized occurrences in ItTenTen and two other corpora of spoken contemporary L1 Italian, the CLIP and the LIP. The CLIP (Corpora e Lessico di Italiano Parlato) is a 342,000-word corpus of 100 hours of spoken Italian divided into five subcorpora (e.g., dialogue, TV broadcast, phone conversations). The LIP (Lessico dell'Italiano Parlato) is a 490,000-word corpus of spoken Italian consisting of 58 hours of monologic and dialogic conversations recorded in five Italian cities in the early 1990s.

L2 data does not reflect the distribution in the Italian input. The results of this study differ from studies supporting AH in that early perfectives in learners' production are not telic, but rather general-purpose, high-frequency verbs like *fare* 'make, do', *dire* 'say', *dare* 'give', *andare* 'go', and *prendere* 'take, get'. Such verbs are actionally underspecified because they can be either telic or atelic depending on the kind of sentential completion. Admittedly, Rastelli (2020a) utilized verbs in isolation – and not predicates – as the unit of analysis because none of the Italian and learner corpora currently available are tagged for the aspectual categories of predicates (at VP level and above). This is an important shortcoming that leaves open the issue of whether, when, and how L2 learners and native speakers alike can derive lexical aspect compositionally by computing the verb and its surroundings.

In an elicited narrative experiment (Rastelli and Vernice 2013), 143 undergraduate American students spending one semester of their second or third university year in Italy for a study-abroad program were asked to describe a short clip in which two telic events overlapped. In the scene used as stimulus, a woman exits a restaurant (event A) and sees her bus with all her travel-mates leaving without her (event B). The ordered sequence is reproduced in Figure 1.

The authors report that, while most native speakers choose the lexically specified telic verb *uscire* 'exit' in the present tense to describe event A (example (2)), about 80% of beginners use the basic verb *andare* in the Italian *passato prossimo* (the perfective past) along with the adverb *fuori* 'outside' perfective, like in sentence (3):

(2) Quando lei esce vede che il suo pullman parte
     'When she exits (she) sees that her bus is leaving'

(3) Quando lei è andata fuori il suo autobus già parte
     'When she went outside, her bus had already left'

These results suggest that beginners – also due to poverty of vocabulary –might express telicity overtly by means of adjuncts expressing the endpoint of the event, rather than encoding it lexically using the target-like verb.[5]

Rastelli (2019) studied the imperfective paradox in an L2 with a dynamic completion-entailment test. The imperfective paradox refers to the fact that the imperfective-progressive yields a completion entailment with atelic verbs (e.g., Livia was pushing the chair → Livia pushed the chair = true) but not with telic verbs (Livia was peeling the tangerine → Livia peeled the tangerine = not necessarily true). The research question asked whether L2 learners are sensitive to the imperfective paradox, just as adult NSs are. Sensitivity to the imperfective paradox is possible

---

[5]An anonymous reviewer observed that the results of the study summarized here do not necessarily provide evidence for the LUH, since one cannot conclude whether beginner students used the basic verb *andare* due to the lack of knowledge of the telic/atelic distinction or because they were simply translating from the English 'went out', which could be used to describe the situation in their mother tongue. This raises the general issue of whether production data can be used to tap into the learners' semantic representations. Nevertheless, a qualitative analysis can identify some possible indicators of the development of learners' aspectual competence, which can be further tested with experimental methods such as those used in psycholinguistic and neuroimaging studies.
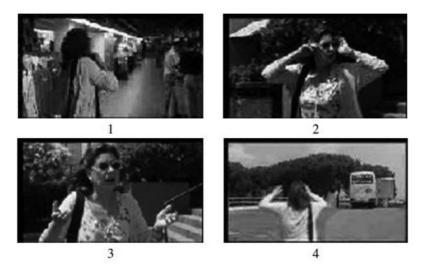
Figure 1: The ordered sequence of frames from the clip used as stimuli. Frames 1 and 2 refer to event A, frames 3 and 4 refer to event B

only if one can distinguish between telic and atelic verbs, as the imperfective paradox works only with the former. A novel technique – the Interval-Based Truth-Value (IBT) judgment test – was utilized in this study. In the IBT, participants watched a short video clip and were instructed to interrupt it by pushing a button as soon as they thought that the person in the video had carried out the action described by the displayed sentence. Each video clip is built around four phases. Figure 2 visualizes how the event described in the perfective sentence with an atelic verb *Livia ha spinto la sedia* 'Livia pushed the chair' flows across the phases of (i) preparation and start (Livia grabs the chair), (ii) duration (Livia pushes it across a room), (iii) culmination (Livia arrives at a desk and stops), and (iv) resulting state (Livia sits on the chair).

Ninety-nine non-native Italian learners at different proficiency levels and with different L1 backgrounds (either Chinese, Russian, or Spanish) took part in this experiment. A total of 32 experimental sentences were derived from eight telic verbs (i.e., accomplishments) and eight atelic verbs (i.e., activities). Each verb occurred once in the perfective and once in the imperfective-progressive. Analysis of reaction times showed that beginner and intermediate L2 learners – unlike native speakers – interrupted the clip regardless of the event completion and did not differentiate between telic and atelic verbs. One may object that the imperfective paradox is not appropriate as a diagnostic for L2 learners, but in the past thirty years, proponents of the AH either relied on their intuitions or used two aspectual diagnostics (i.e., imperfective paradox and adverbial test) to code the telicity of the verbs that learners produced or comprehended. The experiment described in this paper compares Italian native speakers' and L2 Italian learners' responses to an adverbial test.

Figure 2: The event-flow across four phases: start, duration, culmination, and resulting state.

L'uomo blu _____ con l'uomo rosso (*mangiare, parlare, russare, ascoltare*)

4. THE "FOR/IN + TIME SPAN" ADVERBIAL TEST

Dowty (1979: 56) proposed that activity and accomplishment predicates[6] can be distinguished by restrictions on the types of time adverbials they can take and by the entailments they yield when various time adverbial phrases are present: "Whereas accomplishment verbs take adverbial prepositional phrases with *in* but only very marginally take adverbials with *for*, activity verbs allow only the *for*-phrases". Sentences 4a-b and 5a-b (from Dowty 1979) show this contrast.

(4)   a. ?John painted a picture for one hour (acceptable with an iterative meaning: 'J. kept painting a picture over and over')[7]

    b. John painted a picture in one hour

(5)   a. John walked for one hour

    b. *John walked in one hour

An *in* adverbial measures the time span within which eventualities expressed by accomplishments culminate, while a *for* adverbial measures the temporal duration of eventualities denoted by activity verbs (Filip 2012: 722).[8] Not all authors deem the "*for/in* + time span" a reliable test. Some include it in their collection of diagnostics (e.g., De Swart 1998: 350, Slabakova 2001: 23, Borer 2004: 294, Folli and Ramchand 2005; van Valin 2005: 34–42), while others do not. Many SLA studies on aspect do not use aspectual diagnostics at all. I surveyed a sample of 45 SLA studies published from 1998 to 2018 in peer-reviewed international journals,

---

[6]Dowty did not adopt the term 'telicity' in his analysis.

[7]An anonymous reviewer – native speaker of English – observed that sentence (4)a is perfectly fine and does not require an iterative reading, "just a large painting".

[8]For an account of actional shifts induced by modifiers see Filip 2012: 723.

conference proceedings, and edited volumes.[9] In 24 of those studies (about 53%), the authors do not use any diagnostics. Most often, the verbs L2 learners produce/comprehend are simply coded as telic or atelic.[10] Among the 21 SLA studies that utilize aspectual diagnostics, 10 include the adverbial test while 11 do not. The adverbial test systematically conveys two aspectual leaks: A directional PP (e.g., 'to school') in sentence (5)b or a bare plural ('e.g., 'pictures') in sentence (4)a suffice to shift the lexical aspect from atelic to telic and vice versa, respectively, thus rendering English sentences like (6) and (7) acceptable:

(6) John walked <u>to school</u> in one hour[11]   TELIC

(7) John painted <u>pictures</u> for one hour   ATELIC

5. RATIONALE FOR THE STUDY

The adverbial test was designed for linguists and native speakers, not for L2 learners. However, many authors have used this test to code the lexical aspect of the predicates that L2 learners produce or comprehend. To my knowledge, the current study is the first to directly compare native speakers' and L2 learners' performance on the adverbial test. Such comparison is meaningful methodologically (as assessing interlanguage data from the standpoint of the TL may commit the comparative fallacy, see Lardiere 2003) and because both a convergence and a divergence between ratings can be informative. If native speakers' and learners' ratings significantly converge, this can be taken as a cue that they represent lexical aspect alike, as it is often implicitly assumed by some proponents of the AH. If instead the ratings diverge, there could be two explanations, according to the LUH: (i) learners rely on the aspectual distinctions of their L1; or (ii) learners' aspectual competence is under construction. One explanation does not necessarily exclude the other: L1–L2 differences and developmental factors are very likely to interact, thus combining their effects.

---

[9]The sampled journals were *Natural Language & Linguistic Theory, Language Learning, Journal of Child Language, Studies in Second Language Acquisition, The Canadian Modern Language Review, Language, IRAL, TESOL Quarterly, The Modern Language Journal, Second Language Research, Bilingualism: Language and Cognition, International Journal of Bilingualism, Heritage Language Journal, Studia Linguistica, Linguistics, Brain and Language, Language Acquisition, Child Language, Lingua,* and *Memory and Cognition.*

[10]An anonymous reviewer observed that the common practice of assigning lexical aspect to learners' predicates simply by using authors/NSs' intuitions or by measures that are not reported puts at risk the credibility and reliability of any study. I agree. The current paper is not an invitation to avoid using aspectual diagnostics, but to avoid overinterpreting their results when such diagnostics are used with non-native speakers.

[11]To add a culmination point to 'walk' and make the sentence acceptable, the corresponding sentence in Italian would need a heavier PP, like *fino a scuola* 'until school'. The preposition *a* 'to' alone would rather suggest direction (= towards school) and possibly lack of completion.

6. EXPERIMENT

In this section, the research questions of the study are outlined, the material, procedure and participants are described.

6.1 Research questions

There are three research questions (RQs) in this study:

1. RQ_1: Do NSs and L2 Italian learners rate the compatibility of telic and atelic verbs with '*per x*-time' and '*in x*-time' expressions similarly?
2. RQ_2: Do learners' proficiency level and length of immersion affect ratings?
3. RQ_3: Does learners' L1 affect ratings?

6.2 Material

In the current study, telic and atelic verbs were utilized. The label 'telic' in this test applied to both achievements and accomplishment verbs, as they both have a resulting state in their semantic template (section 2). They differ, however, in that achievements are punctual, while accomplishments are durative. It is important to stress that the labels 'telic' and 'atelic' in this study do not perfectly overlap with the categories of 'activity' and 'accomplishment' used by Dowty (1979) in his adverbial test. Indeed, although states and activities are both atelic, Dowty's adverbial test applies only to activities, and not to states. Similarly, although achievements and accomplishments are both telic, they behave differently with respect to the adverbial test. While achievements allow only *in*, and reject *for*, accomplishment verbs are partially acceptable with *for*, at least with an iterative meaning. For example, the experimental sentence used in this study 'he peeled the orange for ten minutes' would mean that 'he peeled the same orange over and over again'. The same iterative interpretation is not possible with the sentence '*he arrived home /*won the race for ten minutes.'

The telic and atelic sentences included in the adverbial test were created following a two-step procedure. The procedure was aimed at establishing a reliable benchmark for comparing NSs and L2 learners. A benchmark for comparison is reliable if (a) NSs' judgments on the lexical aspect of verbs are sufficiently homogeneous and if (b) L2 participants are familiar with the meaning of the experimental verbs, that is, if they can represent at least one event that can be associated with them (see below). Condition (a) is important to avoid or at least minimize the experimenter bias (the observer-expectancy effect) in the choice of experimental verbs (Forster 2000). Condition (b) is important for contrasting the predictions of the AH and the predictions of the LUH. Indeed, The LUH claims that lexical aspect (the telic vs. atelic distinction) is a separate and delayed component of verb meaning. Specifically, acquiring the meaning of a predicate is separate from acquiring its lexical aspect. If L2 learners ignore the meaning of the L2 verbs used in the experiment, one cannot claim that meaning and lexical aspect are separate components that are learned at different points in development.

First, eight telic (accomplishments and achievements) and eight atelic (activity) Italian verbs were selected by the author from three textbooks for beginner and intermediate learners (i.e., levels A2 and B1 of the Common European Framework of Reference [CEFR]).[12] L2 learners' highest possible degree of familiarity with the verbs (or learners' degree of exposure to the verbs) was the criterion guiding the selection. Degree of familiarity/exposure was defined as follows. The verbs were selected from the textbooks used in the Italian language classes that more than 90% of participants of this study were attending (or had attended). In those books, all the selected verbs appeared in the first ten teaching units and they all belonged to the lists of lemmas utilized in the official L2 Italian proficiency test "CILS" (A2–B1 levels)[13] (e.g., Barki et al. 2003: 123–131). This ensured that every instructed L2 Italian learner – whether in Italy or abroad – had likely been exposed to such verbs since the earliest stages of acquisition. Some verbs concerned classroom activities and language learning (*capire* 'understand', *parlare* 'talk', *finire* 'finish'); some other concerned daily activities that learners are often asked to describe (*dormire* 'sleep', *lavorare* 'work', *cantare* 'sing', *camminare* 'walk') and were frequently used in personal narratives. In all accomplishments, telicity was a function of (a) the definiteness of the article (e.g., **il** *romanzo* 'the novel' vs **un** *romanzo* 'a novel'); (b) the cardinality-quantifiability (the degree of incrementality) of the direct object NP (e.g., *sbucciare l'arancia* 'peel the orange'). The interaction of such features always determined telicity compositionally (Verkuyl 1993). Table 1 reports the sixteen experimental verbs:

As the second step, prior to the experiment, all participants took a brief lexical test to check whether they knew the meaning of the selected verbs. The test consisted of sixteen pictures – one for each predicate – depicting an action performed by a blue or red man. Participants were requested to fill in the blank with the proper verb (in any tense) choosing among four options, as in Figure 3:

We tested 365 learners in total, and eventually eliminated 66 learners (18%) who didn't score 100% on the vocabulary test; 299 learners advanced to the judgment task (82% of the initial group, see section 5.4).

Telic and atelic sentences were created based on the sixteen selected verbs (Table 1). Atelic sentences consisted of a human masculine subject (e.g., *Mario*), a verb in third person singular, a direct NP argument or an adjunct, and a '*per* 'for' *x*-time' or '*in* 'in' *x*-time' adverbial. Telic sentences were composed of a human feminine subject (*Maria*), a verb in third person singular, an argument or an adjunct, and a '*per* 'for' *x*-time' or '*in* 'in' *x*-time' adverbial.[14] Each verb occurred twice, in two identical sentences – one with 'in *x*-time', the other with 'per *x*-time' – yielding a total of 32 sentences, 16 acceptable and 16 unacceptable. All verbs in the experiment were

---

[12]<https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>

[13]For an example: <https://cils.unistrasi.it/89/196/Liv._A2_adulti.htm>

[14]An anonymous reviewer pointed out that the fact that all atelic predicates had a masculine grammatical subject while all telic predicates had a feminine subject might have cued participants' responses. It is not possible to determine whether this in fact had any effect.

TELIC		ATELIC	
capire il problema	figure out the problem	*parlare*	talk
vincere la gara	win the race	*dormire*	sleep
finire di scrivere il romanzo	finish writing the novel	*lavorare*	work
arrivare a casa	arrive home	*tossire*	cough
prendere la pillola	take the pill	*ballare*	dance
sbucciare l'arancia	peel the orange	*cantare*	sing
svuotare la borsa	empty the bag	*camminare*	walk
comporre il numero di telefono	dial the telephone number	*ridere*	laugh

Table 1: List of the sixteen Italian verbs used in the adverbial test
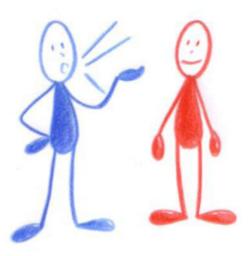

Figure 3: Example of Item in the Lexical Test, *l'uomo blu parla con l'uomo rosso* 'the blue man talks to the red man'

presented in the present tense. There are two reasons for this choice. The first is that the use of the Italian perfective past (the *passato prossimo*) in lieu of the present tense entails the presence of either the auxiliary *avere* 'have' or *essere* 'be' (e.g., *ha dormito* 's/he slept' vs. *è arrivato* 'he arrived'), and the choice between such auxiliaries in Italian is not aspectually neutral, rather it prompts by default an atelic and a telic interpretation, respectively. Indeed, most intransitive monoargumental verbs taking auxiliary *avere* are unergative (agentive and atelic), whereas most intransitive verbs that take *essere* are unaccusative (nonagentive and telic), with about one hundred exceptions to the rule).[15] The second reason for presenting participants

[15]This is true especially with verbs such as *finire*, which can be either unaccusative or unergative, and display double auxiliaries.

with sentences in the present tense is to avoid confounding lexical and grammatical aspects when trying to assess their aspectual competence. Shirai (2013: 288) observed that time adverbials function as perspective-takers of the event, exactly like perfective morphology. Such redundancy would obscure the possibility of teasing apart – in a learner's competence – the knowledge of lexical aspect from the knowledge of grammatical aspect (the perfective vs imperfective morphology). Talmy (1978) wrote that the grammatical element "in – NP – extent of time" specifies event boundedness and – just like the perfective morpheme – focuses automatically on the limits of any event, regardless of the inherent telicity or atelicity of the verb that expresses it[16] (on the issue of separateness between lexical and grammatical aspect, see Filip 2004). Table 2 reports the complete list of the experimental sentences.

The normalized (per million tokens) frequency of verbs in the target input was controlled in ItTenTen, which, to date, is the largest corpus of contemporary Italian.[17] Lemma frequency of verbs – rather than lexeme frequency – was calculated using the function 'concordance' in the SketchEngine software.[18] Frequency of lemmas was meant as a proxy of learners' degree of familiarity with the general meaning of verbs rather than their knowledge of the formal features associated with lexemes (e.g., present vs. past, singular vs. plural etc.). Indeed, the procedure described in this section aimed to minimize the risk that participants did not know the meaning of the verb (their knowledge of its (a)telicity being instead the outcome variable) (condition b). The option of measuring the frequency of each telic predicate as a whole – rather than the frequency of the corresponding lemmas – was also discarded. In fact, the frequency of whole predicates is negligible[19] and its use could have made the comparison between telic verbs and atelic verbs (lacking an object) statistically meaningless. Table 3 reports the normalized frequencies of the 16 verbs used in the experiment across the two lexical aspect categories.

Although telic lemmas are more frequent in the input than atelic ones (on average, atelic lemmas are 50.79 per million tokens less frequent than telic ones), a monofactotorial ANOVA showed that this difference is not significant (df 1, F value = 1.33, p-value = 0.26)

---

[16]As an anonymous reviewer pointed out, this is not a problem with past predicates in English (the language in which the test was created). On the contrary, the test becomes problematic when the Italian present tense is used. In such a case, the English translation of Italian predicates is misleading because in English expressions such as "in one minute" with the present have a future reading, like "the race starts in one minute."

[17]ITTenTen16 is a 4.9-billion-word web corpus (downloaded by SpiderLing from May to August 2016) made up of texts collected from the Internet. The corpus is a part of the TenTen corpus family, a set of web corpora built using the same method.

[18]<https://sketchengine.eu>

[19]For example, *prendere la pillola* 'take the pill' has 0.87 occurrences per million tokens, *svuotare la borsa* 'empty the bag' has 0.02 occurrences, *comporre il numero* 'dial the number' has 0.55, and *sbucciare l'arancia* has 0.08 occurrences per million tokens.

TELIC sentences		ATELIC sentences	
*Maria capisce il problema in/*per dieci secondi*	Maria figures out the problem in/ *for ten seconds	*Mario parla con suo fratello per/*in cinque minuti*	Mario talks to his brother for/*in five minutes
*Maria vince la gara in /*per dieci secondi*	Maria wins the race in/*for ten seconds	*Mario dorme profondamente per/*in un'ora*	Mario sleeps deeply for/*in one hour
*Maria finisce di scrivere il romanzo in/ *per un mese*	Maria finishes writing the novel in/*for one months	*Mario lavora ogni giorno per/*in otto ore*	Mario works every day for/*in eight hours
*Maria arriva a casa in/*per cinque minuti*	Maria arrives home in/*for five minutes	*Mario tossisce continuamente per / *in in un'ora*	Mario coughs constantly for/*in one hour
*Maria prende la pillola in/*per pochi secondi*	Maria takes the pill in/*for a few seconds	*Mario balla sulla pista per/*in un'ora*	Mario dances on the floor for/*in one hour
*Maria sbuccia l'arancia in/*per dieci secondi*	Maria peels the orange in /*for ten seconds	*Mario canta a voce alta per/*in mezz'ora*	Mario sings aloud for /*in half an hour
*Maria svuota la borsa in/*per un minuto*	Maria empties the bag in/*for one minute	*Mario cammina per strada per/*in dieci minuti*	Mario walks along the road for /*in ten minutes
*Maria compone il numero di telefono in/*per cinque secondi*	Maria dials the number in /*for five seconds	*Mario ride con sua moglie per/*in cinque minuti*	Mario laughs with his wife for/*in five minutes

Table 2: The experimental sentences

6.3 Procedure

This experiment used a prompted acceptability judgment task. Sentences were presented on a white screen of a PC monitor following a fixation cross. After six seconds, an acoustic signal prompted participants to rate the sentence on a pre-formatted 10-point scale on a sheet of paper. Participants had five seconds to rate each sentence before another acoustic signal occurred and a white blank screen with the fixation cross appeared. A warmup trial was run to allow participants to familiarize themselves with the timing. Experimental sessions lasted approximately 34 minutes in total, including preliminary examples and warmup. Participants could not return to sentences once they had been rated. The task wording was the following: "We want you to tell us for each sentence whether you think it sounds natural in Italian. A high rate means that the sentence sounds very natural to you. A middle rate means that you feel that there is something odd/wrong with the sentence and that you would need to think about it before saying it. A low rate means that the sentence

does not sound natural to you." In the majority of cases (e.g., Chinese, Spanish, German, Russian), task wording and instructions were given in the participant's native language. In all other cases instructions were given in both Italian and English. The word 'grammatical' was carefully avoided in both the explanations and the task wording. The following options on the scale were discussed with participants: 10–9 = Definitely natural; 8–7 = Natural; 6–5 = Probably natural; 4–3 = Probably unnatural; 2–0 = Definitely unnatural. Scores were averaged and regressed in statistics considered by their exact values and were not conflated with any adjacent category. The experimental sentences were interspersed with 32 fillers (half correct and half incorrect) that targeted different phenomena of Italian morphosyntax (e.g., number and gender agreement in the NP and VP, clitic pronouns, etc.). A pseudo-randomized order was utilized to ensure a proper distance between two occurrences of the same predicate. Each group of participants rated the same sentences but in a different, pseudo-randomized order. This experiment setup was meant to ensure that (a) each participant saw multiple items per condition; (b) items formed minimal pairs; (c) lexical repetition, adjacency effects, and familiarization with the task were avoided; and (d) the various conditions within the experiment were disguised.

6.4   Participants

A sample of 299 adult (age range 18–29 years; mean = 21.3), non-native speakers (NNSs) of Italian took part in the study. Participants were Erasmus students, international exchange students, and Marco Polo Turandot[20] students from six Italian universities located in North and Central Italy. All participants had normal or corrected-to-normal vision, and 270 of them were right-handed. Participants were informed about the general aim of the experiment, but not about the target structures, and expressed written consent (adapted from Mackey and Gass 2005: 322) to participate in the study. They were all volunteers and were not paid. In order to establish proficiency levels, a CILS (*Certificato di Italiano Lingua Straniera* 'Certificate of Italian as a foreign language') test was administered to all participants before the beginning of the experiment. The CILS test is an official L2/FL Italian proficiency test designed by the University for Foreigners of Siena. The test adheres to the guidelines of the Common European Framework of Reference for languages.[21] Based on an estimate of learners' proficiency, this experiment used the B1 level test, which is composed of four sections: listening comprehension, reading comprehension, metalinguistic knowledge, and written composition. Following an established practice at Italian Universities, students who scored between 90 and 100 (M = 89.5, SD = 6) were considered advanced (C1–C2); students who scored from 60 to 89 (M = 71.5, SD = 9.6) were considered intermediate (B1–B2); students who scored less than

---

[20]The Marco Polo Turandot is a Chinese-Italian bilateral agreement that allows Chinese prospective students to come to Italy for one year prior to enrolling in an Italian university in order to study the language. For more details and information about students, see <http://marcopolo.unipv.eu/>.

[21]See samples of exams at <https://cils.unistrasi.it/89/188/Esempi_di_prove_di_esami.htm>.

LEMMA	Lexical Aspect	n. of occurrences in ItTenTen (per million tokens)
parlare 'talk'	atelic	494.09
dormire 'sleep'	atelic	44.33
lavorare 'work'	atelic	244.38
tossire 'cough'	atelic	0.91
ballare 'dance'	atelic	14.26
cantare 'sing'	atelic	46.7
camminare 'walk'	atelic	40.2
ridere 'laugh'	atelic	37.79
MEAN FREQ		**115.33**
capire 'grasp'	telic	287.95
vincere 'win'	telic	144.01
finire 'finish'	telic	192.34
arrivare 'arrive'	telic	458.31
prendere 'take'	telic	477.95
sbucciare 'peel'	telic	3.65
svuotare 'empty'	telic	9.95
comporre 'dial'	telic	161.06
MEAN FREQ		**216.9**

Table 3: Normalized frequencies of experimental verbs (lemmas)

60 (M = 40.5, SD = 8.6) were considered beginners (A2).[22] In order to get a more refined subdivision and to ensure symmetry among levels, we divided beginning learners into 'absolute beginners' (A1) and 'false beginners' (A2) depending on the score on the proficiency test. Table 4 shows that the distributions of L2 participants across proficiency levels were well balanced.

Information about immersion, operationalized as participants' length of stay in Italy (expressed in days), was collected through a written questionnaire prior to the experiment and to the lexical test. Fifteen different L1s were represented in the learner group, as shown in Table 5.

In cases of bilingual and quasi-bilingual speakers (Spanish–Catalan, Russian–Ukrainian), participants were coded as L1 speakers of a language depending on the language spoken at home and/or with close relatives and friends. Given that the research

---

[22]Rather than unique scale tests like the one adopted in this study, other authors prefer differentiated tests for each L2 proficiency level. Unique scale (or 'curve' methodology) is adopted in many experimental studies, especially in those that regress proficiency onto online measures (e.g., reaction times). Unique scale tests have the advantage of making standard deviations (SD) directly comparable. Through SD comparison, researchers can see whether experimental groups are homogeneous. SD comparison is not possible if one uses centring and standardization of scores that come from different tests (e.g., z-scores). The choice of cutoff points between levels adopted in this study followed the standard practice of placement tests commonly adopted by the Italian universities where the experiment took place.

proficiency	A1	A2	B1	B2	C1	C2
participants	45	50	55	49	51	49

Table 4: Distribution of participants across proficiency levels

English	24	Russian	11
Catalan	10	Serbian	1
Chinese	111	Slovak	3
Czech	1	Slovenian	2
French	4	Spanish	81
German	16	Ukrainian	2
Polish	16		
Portuguese	17		

Table 5: Distribution of the 299 learners across the 14 L1s

questions investigate the separate impacts of proficiency and L1, one must be sure that these independent variables are not nested. In fact, a Kruskal-Wallis test confirmed that there was no interaction between participants' L1 belonging to a given language family and participants' proficiency levels (Kruskal-Wallis chi-squared = 112.17, df = 3, p-value = .15). Table 6 shows that the distribution of the L2 learners across typological families and proficiency levels is in fact well-balanced.

Finally, a sample of 91 NS controls (mean age = 22) was recruited among Italian undergraduate students at the University of Pavia. None of them was a student of linguistics.

7. ANALYSIS

Non-rated sentences (2% of the total) were excluded from the analysis. Such items did not cluster significantly and were evenly distributed across the relevant categories: 55% of them occurred in acceptable combinations, and 45% in unacceptable combinations, 55% were telic verbs whereas 45% were atelic verbs. The difference between telic verbs taking a direct object (e.g., *prendere la pillola* 'take the pill') and the only telic verb having a sentential completion (*finire di scrivere il romanzo* 'finish writing the novel') was taken into account in the statistical analysis, as well as the virtually different telicizing effect of different delimiting objects (e.g., *la borsa* 'the bag' vs. *il numero di telefono* 'the phone number'). Since the data were normally distributed according to a Bartlett test (all *p*-values ≥ 0.6), mono and multi-factorial ANOVAs were used for the analysis (with software R, version 3.5.1, package 'car', R Core Team 2015). Effect-size was calculated via eta-square (η^2). The only dependent variable of the study was the participants' ratings. Independent variables included: learners' L1, typological family, levels of

	A1	A2	B1	B2	C1	C2	**Total**
Germanic	3	9	8	9	7	4	**40**
Romance	16	15	25	15	22	18	**112**
Sino-Tibetan	20	20	20	18	20	13	**111**
Slavic	6	6	6	7	7	4	**36**
Total	**45**	**50**	**55**	**49**	**51**	**49**	

Table 6: Distribution of NNS participants across proficiency levels

	*atelic+in		atelic+per		*telic+per		telic+in	
	M	SD	M	SD	M	SD	M	SD
L2 learners	4.99	3.7	7.05	2.7	5.53	3.5	6.4	2.9
Native speakers	1.7	1.3	8.4	1.6	0.6	1.4	8.5	1.4

Table 7: Means (M) and standard deviations (SD) of ratings across participants and conditions

proficiency, and length of immersion in Italy. Frequency of lemmas entered the model as a covariate (see Table 3).

8. RESULTS

In this section, the results of the study and descriptive and inferential statistics are presented.

8.1 Between-groups comparison

L2 learners and NSs rated the compatibility of telic and atelic verbs with time adverbials very differently. This difference was highly significant according to a monofactorial ANOVA (df = 1, F = 133.22, p value***, $\eta^2 = 0.84$). Table 7 indicates that the standard deviation (SD) was similar within groups. NSs' ratings were especially homogeneous, confirming that the choice of Italian verbs met the purposes of the experiment.

A multifactorial ANOVA showed that NSs and L2 learners evaluated both the unacceptable (df = 1, F = 849, p value***, $\eta^2 = 0.76$) and acceptable combinations (df = 1, F = 228, p value***, $\eta^2 = 0.81$) differently, even though the contrast between groups is much more evident in the former condition (see the left-side panel in Figure 4). Ratings of unacceptable combinations also significantly differed: the *atelic + *in* combination was rated lower than the *telic + *per* combination (mean rating 4.99 vs 5.53 respectively) (df 1, F-value =14.93, p-value***, $\eta^2 = 0.38$).
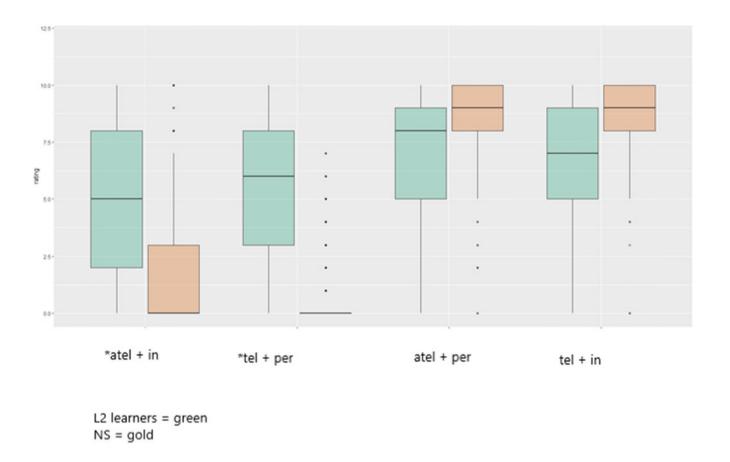
L2 learners = green
NS = gold

Figure 4: Mean of NSs' and L2 learners' ratings across unacceptable and acceptable combinations

The multifactorial ANOVA also showed that NSs rated unacceptable telic combinations significantly lower than unacceptable atelic combinations (M = 0.64 and M = 1.72, respectively). The same differences were not found in NSs' ratings of acceptable combinations. L2 learners, on the other hand, rated all unacceptable combinations similarly, regardless of lexical aspect (F = 298, *p*-value***, $\eta^2 = 0.54$). Finally, a Tukey post-hoc test revealed that there were no significant differences between NSs' and L2 learners' ratings of telic verbs depending on the telicizing NP (e.g., *svuota la borsa* '(she) empties her bag' vs *vince la gara* '(she) wins the race'). The ANOVA also showed that there was no significant difference between the four stimuli featuring achievement verbs and the four stimuli featuring accomplishment verbs (*p*-value = 0.21) and among the telic verbs taking a direct object (e.g., *prendere la pillola* 'take the pill') and the only telic verb having a sentential completion (*finire di scrivere il romanzo* 'finish writing the novel") (*p*-value = 0.68).

8.2   Learners' proficiency

L2 learners with different proficiency levels were differentially sensitive to the adverbial test. Table 8 shows L2 learners' mean ratings of acceptable and unacceptable combinations across proficiency levels.

At a glance, Table 8 and Figure 5 suggest that proficiency strongly affected learners' ratings of unacceptable combinations but only marginally affected ratings of acceptable ones. The outcome of a multifactorial ANOVA analysis confirmed this impression (F = 169, *p*-value***, $\eta^2 = 0.61$).

Both Table 8 and Figure 5 also show the existence of a considerable gap between B2 and B1 learners in the rejection of unacceptable combinations, with the former significantly outperforming the latter. This gap may indicate that proficiency is not a continuous variable, and that the cut-off points separating proficiency levels should be taken with caution. These are categorical variables which – at least to some extent – might also depend on the nature of the test and on some preliminary decisions on how L2 competence should be tested. In fact, if one collapses the six proficiency levels into three macro-categories (beginner, intermediate, advanced) one sees that a more gradual progression is restored (mean ratings of unacceptable combinations: beginner = 5.87; intermediate = 4.79; advanced = 3.71).

proficiency levels	unacceptable combinations	acceptable combinations
A1	6.59	7.08
A2	5.71	6.83
B1	5.41	6.74
B2	3.67	6.41
C1	3.81	7.93
C2	3.66	6.87

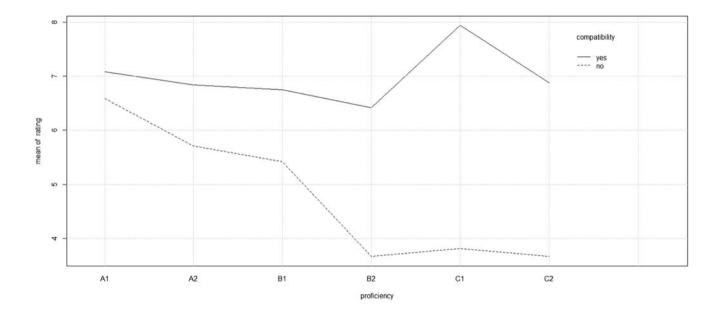**Table 8:** L2 learners' mean ratings across proficiency levels

Figure 5: L2 learners' mean ratings of unacceptable and acceptable combinations across proficiency levels
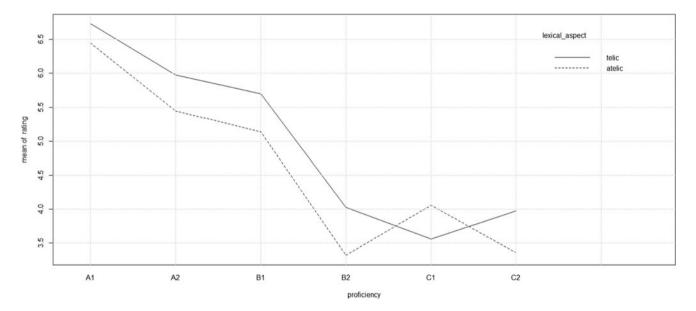
Figure 6: L2 learners' mean ratings of unacceptable and acceptable combinations across telic and atelic conditions

L2 learners rated unacceptable combinations involving both telic and atelic verbs similarly across all proficiency levels, as shown by Figure 6.

Even the most advanced learners' mean ratings of unacceptable combinations were significantly higher than NSs' (M = 3.71 and M = 1.18, respectively) (df = 1661, F = 2062, p-value***, $\eta^2 = 0.78$).

Finally, as we have already seen, ratings of unacceptable combinations significantly differed between *atelic + *in* and *telic + *per* (mean rating 4.99 vs 5.53 respectively). This difference was not affected by learners' increasing proficiency (p-value = 0.74).

8.3   Learners' L1

L2 learners with different L1s were differently sensitive to the adverbial test. Table 9 reports the mean and the standard deviation (SD) of participants' ratings of acceptable and unacceptable combinations across the four learners' L1 typological families:

In general, learners' ratings of acceptable combinations were far more homogeneous than the ratings of unacceptable ones. In particular, the ANOVA and Tukey's *post-hoc* test showed that in the 'acceptable' condition, no between-groups difference reached significance (with Slavic and Romance groups patterning alike also in the unacceptable condition). In contrast, all between-groups differences in the 'unacceptable' conditions were significant (p-values ***). Standard deviations were quite similar among groups and across conditions, except for Sino-Tibetan learners. As for between-language comparison, Table 10 suggests that learners with Germanic L1 (English and German) rated unacceptable combinations more similarly to Italian NSs than speakers of other Romance languages. The difference between languages was significant overall (df 3, F-value = 5.308, p-value = 0.0012, η2 = 0.48), but Tukey's *post-hoc* suggested that only the means of the differences between Spanish–English, Portuguese–English and Portuguese–German reached statistical significance (p-value ≤ 0.05):

It should be observed, however, that between-language differences are likely due to the differences in the composition of the learner sample: almost 50% of Spanish learners were beginners or absolute beginners, whereas this percentage goes down to 20% for L1 English learners. A multifactorial ANOVA confirmed that the interaction

Language family	unacceptable combinations		acceptable combinations	
	mean	SD	mean	SD
Native speakers	1.18	2.03	8.48	1.55
Germanic	3.67	3.03	6.49	3.12
Romance	4.68	2.86	6.52	2.71
Slavic	5.05	3.32	6.99	2.88
Sino-Tibetan	6.44	4.31	7.06	2.92

Table 9: Learners' ratings across typological families

Language	unacceptable combinations		acceptable combinations	
	mean	SD	mean	SD
Native speakers	1.18	2.03	8.48	1.55
English	3.7	2.83	6.31	3.13
German	3.57	3.33	6.82	3.17
Portuguese	5.22	2.37	6.41	2.64
Spanish	4.63	2.89	6.49	2.62

Table 10: Ratings by learners speaking Germanic and Romance languages (French and Catalan were excluded from this analysis due to scarcity of data points)

between L1 and learners' proficiency levels significantly modulated the ratings (df 8, F-value = 4.81, p-value***, η2 = 0.45). A Tukey's post-hoc revealed that this interaction was stronger for lower proficiency levels (A1, A2) (p-value = 0.005) and was not significant at higher levels (p-value = 0.26). As to the interaction between proficiency levels and L1, among beginning L2 learners, those speaking Germanic and Romance languages rated unacceptable combinations more similarly to NSs, whereas Slavic and especially Chinese learners' ratings deviated. Differences among these groups were all highly significant in the unacceptable condition (DF = 3706, F = 323, p-value***, $\eta^2 = 0.57$) but not in the acceptable one (p-value = 0.55), as shown in Figure 7.

Intermediate L2 learners with Germanic, Slavic, and Romance L1s (in this order) were significantly closer in their ratings to NSs than L1 Chinese learners, as shown in Figure 8. Again, the difference between the L1 Chinese speakers and learners with other L1s was highly significant in the unacceptable condition (DF = 3914, F = 317, p-value***, $\eta^2 = 0.68$) but not in the acceptable condition.

Finally, there was a significant interaction between L1 and type of unacceptable combinations (Df, 4, F-value = 10.26, p-value***). Table 11 reports the ratings of *atelic + in* and *telic + per* combinations across typological families:

Both between-family and within-family differences in the ratings of unacceptable combinations were significant, except for the Romance group. None of the differences were modulated by learners' proficiency (p-value = 0.44).

8.4 Length of immersion

Learners' length of stay in Italy at the time of the experiment did not affect ratings (p-value = 0.175), and only weakly correlated with learners' proficiency (Pearson correlation = 0.30, t = 23.501, df = 5390, p-value***).

9. DISCUSSION

In sum, regardless of their L1s, the majority of beginner and intermediate L2 Italian learners – unlike Italian NSs – rated the 16 unacceptable combinations of frequent,
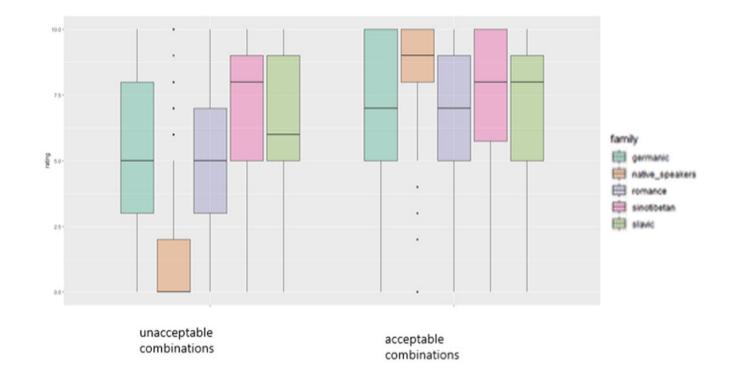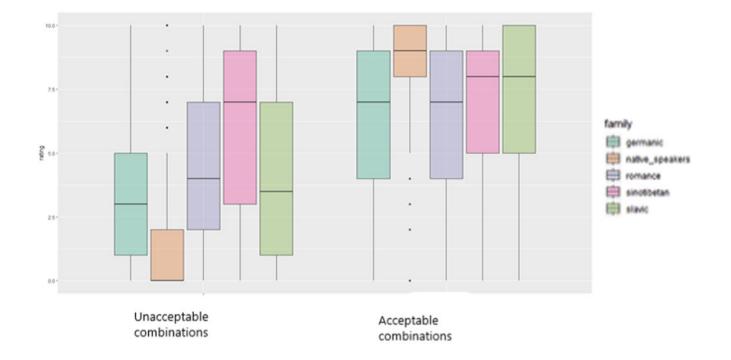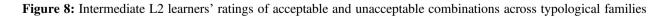
Figure 7: Beginner L2 learners' mean ratings of acceptable (right) and unacceptable (left) combinations across typological families

Figure 8: Intermediate L2 learners' ratings of acceptable and unacceptable combinations across typological families

Language family	*atelic + *in*	*telic + *per*
Native speakers	1.72	0.64
Germanic	3.23	4.11
Romance	4.50	4.86
Sino-Tibetan	6.06	6.82
Slavic	4.75	5.35

Table 11: Learners' mean ratings of unacceptable combinations by typological family

familiar telic and atelic verbs with time adverbials as either 'probably natural' or 'natural'. Lexical aspect (telic vs. atelic), and type of telicizing NP did not affect learners' ratings. L2 learners accepted the 16 acceptable combinations as 'natural', but their ratings were not as high as those of the NSs. The gap between L2 learners and NSs diminished as learners' proficiency increased, although it did not disappear completely. In fact, even the most advanced (C2) learners tended to rate as 'probably unnatural' those combinations NSs rated as 'definitely unnatural'. Proficiency affected learners' ratings of unacceptable combinations but not of acceptable ones. Learners' L1s also affected participants' ratings of unacceptable combinations, often in interaction with proficiency. Slavic and Chinese learners showed the least nativelike performance at initial stages of acquisition. However, at higher proficiency levels, Slavic learners' performance was comparable to that of the other learners. Finally, learners' length of immersion did not affect ratings.

9.1 Is it the test?

The adverbial test – like all telicity diagnostics – was not designed for learners, but for linguists, especially semanticists. Even native speakers who are unfamiliar with lexical aspect categories may have difficulty identifying them without training. Indeed, the reason for adopting the test in this study was not to judge the accuracy of the Vendlerian categories, but to question researchers' implicit assumption that 'telic' for linguists and 'telic' for the learners coincide. Our results indicate that they likely do not. Although the reasons for such a discrepancy may be unclear (see below), its very existence should suggest caution when assessing learners' aspectual competence. It could be objected that the adverbial test is not suitable as a benchmark for either native speaker's and learner's knowledge of aspectual categories. Undoubtedly, test sentences have confounds, including frequency adverbs, manner adverbs, and prepositional phrases which learners may be evaluating, rather than the verbs, when they rate sentences. Prepositions are also notoriously tricky for L2 learners. A real acquisitional challenge concerns the (across- and within-language) polysemy of prepositions *in* and *for*. For example, in Italian, the crucial prepositions *in* 'in' and *per* 'for' are both extremely frequent and polysemous: *in* serves to locate events not only in time but also in space (*vivo in Italia*, 'I live in Italy'), and *per* is also used to express purpose and cause. Since in the adverbial test,

participants' knowledge of prepositions *in* and *per* is essential, L2 learners' proficiency could explain much of the variance in the data. Perhaps a different telicity diagnostic, based on visual stimuli, without adverbs and prepositions, could have worked better. This is exactly what Rastelli (2019) did with the imperfective paradox (section 3.4), which is possibly the most widely utilized diagnostic in the AH studies. Rastelli (2019) uses an interpretation task based on the completion entailment diagnostics and the results showed that L2 learners did not differentiate between telic and atelic verbs. Since two studies – based on two different diagnostics – gave comparable results, one can infer that the problem is not the type of test, but the existence of profound differences in how NSs and beginning learners represent aspectual categories.

9.2   Is it the L1?

The adverbial test seems to have worked differently depending on the L1. Participants who deviated more significantly from native speaker' ratings were Slavic and Chinese learners. In contrast, learners with Germanic and Romance L1s patterned more similarly to native speakers, even if their ratings were not homogeneous. We recall that four typological families were represented in our participant sample: Romance (French, Catalan, Spanish and Portuguese), Slavic (Czech, Polish, Russian, Serbian, Slovak, Slovenian, Ukrainian), Sino-Tibetan (Mandarin Chinese), and Germanic (German and English). A source of difficulty for learners with Slavic L1s could be that in all Slavic languages, preverbal morphemes indicate perfective aspect, while null suffixes indicate that an event is ongoing (Slabakova 2001: 5). Some authors maintain that these morphemes conflate [+telic] and [+perfective] values, which therefore cannot easily be disentangled. In contrast, in the Romance family, grammatical aspect is encoded overtly through verb morphology, while lexical aspect is covert (Montrul 2002: 42). In Mandarin Chinese, aspect is coded by four free morphemes which do not express tenses, but rather different perspectives on the situation. Particles *zhe* and *zai* signal that situation is imperfective, progressive, or durative, whereas *le* and *guo* express perfective aspect (Li and Thompson 1981). It is not entirely clear whether aspectual markers in Chinese encode only grammatical aspect or also its interaction with lexical aspect.

    Another source of divergent behaviours among learners could be the opposition between Germanic and Slavic languages (see Filip 2004, and Ayoun and Rothman 2013 for a review). Two different sources of telicity are relevant when comparing those families: a dedicated functional projection above the VP (called AspP) and the presence of a telicity feature (quantifiers, definite article, and accusative case) in the direct object. In Germanic languages, the presence of a definite article or the expression of a defined quantity (as opposed to bare plurals) will trigger a telic interpretation. In Slavic languages, verbal morphology triggers the telic interpretation, since a prefix can cause an atelic predicate to become telic. However, extensive research on L1 acquisition has shown that telicity could be easier to learn when it is overtly marked (as in Slavic languages) rather than when it must be computed from the properties of the VP and its object together (as in Germanic languages) (van Hout 2008). In this respect, one should also consider that typological families

are not monolithic as to how they encode either lexical or grammatical aspect. For example, within the Slavic family, Bulgarian has articles, while Russian and Czech do not. This may affect learners' sensitivity to the existence of a delimiting feature in the DP (Di Sciullo and Slabakova 2005).

In general, the results of this study do not allow generalizations that are strictly based on L1–L2 similarities. First of all, L1 – along with proficiency – had an impact only on the ratings of unacceptable combinations, not on acceptable ones (see below). Second, it was not always possible to disentangle the impact of L1 from the impact of proficiency, partly because the sample was not perfectly balanced across proficiency levels (Table 5). For example, most L1 German participants were intermediate-advanced, while most L1 Chinese participants were beginner learners. As we have seen, some apparently puzzling findings can be explained in terms of sample composition. German learners rated acceptable and unacceptable combinations more similarly to Italian NSs than L1 Spanish learners did, likely because German learners were, on average, more proficient than L1 Spanish learners.

There are also cases in which the impact of the L1 seemed completely independent of the sample composition and *not* modulated by proficiency. For example, the remarkable differences in how learners rated the *atelic + *in* vs *telic + *per* unacceptable conditions (Table 10) significantly depended on L1 (typological families) only. However, this did not hold for all learners. For example, L1 Chinese learners of Italian in our sample rated acceptable and unacceptable combinations very similarly, albeit standard deviation – and therefore the amount of variance in responses – was much higher in the latter condition. Yet, if one looks only at most proficient Chinese learners (level C2), they rated unacceptable and acceptable combinations very similarly to advanced learners of all other groups (mean for unacceptable combinations = 3.57; mean for acceptable combinations = 6.85). If typological distance between Italian and Chinese impacted on ratings, this impact was certainly limited to beginner and lower-intermediate learners. In sum, if one is looking for the possible effect of learners' L1 in isolation, this seems restricted to the differences in the ratings of unacceptable combinations by some groups of beginning learners.

Another fact that needs to be explained is why all learners – regardless of their L1 – found it much more difficult to reject incorrect sentences than to accept correct ones. This has also been generally observed in previous L2 studies on aspect. Nishi and Shirai (2019) argued that L2 learners have difficulties in rejecting incorrect L2 aspectual structures (but not in accepting correct ones) when such structures involve an L1–L2 discrepancy in lexical aspect. The authors claimed that their results confirm a strong L1 effect at the level of surface inflected verbal forms, showing significantly higher accuracy for items for which direct translation yields correct meaning than those that do not. In our data, acceptance of correct sentences was actually quite good regardless of L1 and proficiency level, while rejection of incorrect sentences only obtained at higher levels of proficiency, partly regardless of learners' L1. Therefore, it may be argued that in the present study, such asymmetry did not come from L1–L2 differences or (lack of) knowledge of telicity per se, but from more general tendencies that were observed in other SLA studies that used acceptability judgments, not only in the domain of aspect (Plonsky et al. 2019). If

accepting correct combinations is easier for L2 learners than rejecting incorrect ones, regardless of the L1, this, of course, may have affected the results of this study as well. If beginning and intermediate learners adopt acceptance as default when they do not know enough of the L2, then the acceptability judgment technique should perhaps be confined to subjects at high(er) proficiency levels.

9.3   Is it second language development?

There is also a developmental explanation for the results of this study, namely that L2 telicity is under construction, alongside the whole tense-aspect system, as predicted by the LUH (section 3). Since re-construction of L2 aspectual categories takes time, it does not come as a surprise that proficient learners performed significantly better than beginning learners in *all* telicity tests investigated so far, not only the adverbial test. The LUH claims that a developmental pattern exists, constraining the ways in which learners can represent the lexical aspect of L2 verbs over time, regardless of their L1s. Beginning L2 learners – unlike native speakers – may initially be uncertain about the telicity of verbs. The learning algorithm of telicity would stipulate that in a beginner-to-intermediate learner's competence, the telic vs. atelic distinction may be left underspecified. In the meantime, learners focus on features of verb meaning other than lexical aspect. As their proficiency increases, learners gradually learn to recognize the features distinguishing telic from atelic verbs.

An issue that remains largely unaddressed is the nature of learners' incomplete aspectual knowledge and the counterintuitive idea that verbs – at least for a certain period – can be comprehended and used for their general meaning, regardless of lexical aspect. How is it possible that verb meaning and lexical aspect dissociate? To what extent do learners know the meaning of a verb without knowing its aspectual characteristics? For example, can a learner be credited to know what 'run' means in English if they always connect this verb to inherently terminative events such as 'run across a street'? The hypothesis that telicity is learned does not imply that learners' competence does not include the telic vs. atelic contrast. Yet, learners are likely to inherit this distinction from their L1. Rather, the LUH claims that beginning learners may ignore how this distinction is encoded in the L2 verbal lexicon, and, as a consequence, in a beginning learner's competence, one could expect that a single verb expressing motion such as *andare* 'go' could cover all kinds of motion events (directed, undirected, manner of motions, deictic, etc.). Similarly, in the learner's interlanguage, one single verb of perception (via sight) could cover the meanings of 'see', 'watch', and 'observe', or the distinction between 'talk', 'tell', and 'say' could be blurred. Of course, the results of the present study cannot be conclusive about the LUH or about the claim that lexical aspect as well as the whole tense-aspect system is learned. The improvement in ratings may imply that our L2 learners started with underspecified aspectual values and eventually learned the telicity of L2 predicates. However, the same improved ratings could also indicate learners' improvement in overall proficiency, which in turn would have allowed them to judge the acceptability of constructions at the surface level, with the lexical aspect of L2 predicates being still underspecified. Such improvement in learners' overall

proficiency might have especially reflected the temporal frame of the experimental stimuli, that is, the value of Italian prepositions *in* and *per* and the meaning of some temporal expressions (e.g., *quando* 'when'). As we have seen, temporal expressions and prepositions are difficult for beginners to master. Since the test did not include prepositions and time adverbials (section 5.4), it is impossible to tease apart learners' knowledge of L2 lexical aspect from other non-verbal – albeit aspect-related – features of the L2 competence.

The current study has its limitations. Possibly the most critical is that offline tests such as untimed acceptability judgments – unlike timed tasks – cannot tap into an L2 learner's implicit representations (e.g., Jegersky and VanPatten 2013). It is believed that offline tests can (at most) bring to light the effect of learners' reasoning on language or the reflex of declarative knowledge of grammar rules taught in the classroom. Nevertheless, there are two arguments in favour of utilizing offline tests in order to tap L2 learners' knowledge of lexical aspect. First, to my knowledge, it is very unlikely that lexical aspect is explicitly taught to L2 Italian learners in the classroom. Unlike grammatical aspect (e.g., Samu 2020), the very notion of lexical aspect is absent from textbooks, syllabi, teaching materials, and proficiency tests. It is, therefore, very unlikely that participants' reflections on lexical aspect could have been mediated by declarative knowledge of explicit notions rather than by mere intuition. If L2 learner's and NS's ratings diverged, this was probably because of the different nature of underlying representations, and not because of declarative knowledge. Second, under certain circumstances, offline measures may even present advantages for researchers. Without the time pressure typically imposed by tasks involving reaction times, learners in our study might have had more time to carefully consider not only the meaning of the verb but also all the elements in its syntactic surrounding and the pragmatics. These are precisely the factors we would expect to co-determine participants' responses to an adverbial test.

10. CONCLUSION: OPEN ISSUES, CONNECTIONS WITH OTHER THEORIES, AND THE NEED FOR NEW DATA

The LUH claims that beginning learners might take some time to represent the lexical aspect of L2 predicates because this is a separate and perhaps delayed component of the verb meaning. The findings of the current study support such a hypothesis. However, there are also open issues and questions that remain unanswered, the first being why learners' reliance on information about lexical aspect would be temporarily severed or suspended. Why would learners block access to a source of information that is central to the representation of temporality in the L2? One possible answer is that temporarily disabling aspectual knowledge somehow simplifies the learner's task and spares processing resources, which are already drained. Yet, the LUH claims that what is temporarily disabled is not the representation of aspect or temporality (which can be deduced from a learner's L1), but a learner's capacity to map it onto the novel L2 lexicon, which is a costly, time-consuming and perhaps also delayable ability. After all, a beginning learner of Italian may well

use the basic and underspecified verb *andare* 'go' instead of more aspectually speci-fied verbs such as *venire* 'come', *entrare* 'enter', or *arrivare* 'arrive' without the risk of being misunderstood by interlocutors. Indeed, using L2 predicates for their general meaning – and disregarding whether they are telic or atelic – is, in many situations, simpler and more straightforward. Such a strategy may parallel the 'one meaning one form' principle that has been proposed to account L2 acquisition of functional morph-ology (Slobin 1979, Andersen 1984), according to which beginners in particular would tend to avoid polysemy and multifunctionality and would prefer mapping one meaning onto just one lexical entry (and vice versa). If the learning algorithm stipulates that in a beginning-to-intermediate learner's competence, the telic vs. atelic distinction of L2 predicates is underspecified, learners would focus on features of the verb meaning (e.g., formal features such as number, person etc.) rather than on lexical aspect.

The second, somewhat unexplored, issue is that L1 – like proficiency – had an impact at least on the ratings of unacceptable combinations. This could indicate that that lexical aspect is not being ignored under all conditions, and that when the L1 tends to match the target (e.g., in Romance languages), learners make use of that source of information. Future research should investigate the conditions that lead learners to put more or less weight on the lexical aspect of predicates. However, the results of the current study seem to suggest that the effect of learners' L1 is restricted to some groups of beginning learners and not to all participants. As noted above, proficiency seems more important than L1 as an explaining variable.

The LUH is connected with at least one other aspectual theory. The novelty of the LUH resides in the idea that L2 lexical aspect, along with all the other elements in the Tense–Aspect system, are under construction and therefore must be learned. The idea that the information relative to L2 lexical aspect might not be fully available to beginners is not new. The Default Past Tense Hypothesis (DPTH, see Salaberry 2008, 2011: 184) significantly remodulated the tenets of the AH and introduced the factor of L2 proficiency into the picture. The DPTH predicted that beginners use only one default form to express all past meanings. This form is more likely to be the perfective than the imperfective because the latter is cognitively more complex, semantically more subtle, and crosslinguistically less uniform and less fre-quently attested. Most important, the DPTH predicts that lexical aspect (along with discourse grounding) increasingly affects past tense marking, as learners gain more experience with the TL. In other words, L2 learners would not selectively associate lexical aspect and aspectual morpheme(s) from scratch. On the contrary, they would move toward prototypical (telic-perfective vs. atelic-imperfective) associations grad-ually, as their knowledge of the language increases. Salaberry (2011) argues that as non-native speakers gain more experience with the TL, they may develop an increas-ingly accurate system of proceduralized knowledge based on probabilistic frequen-cies associated with lexical aspectual values and, to some extent, on discourse grounding associated with an expanded scope of information (as provided by view-point aspect). Therefore, the claims of the LUH may be relevant for intermediate/advanced stages of acquisition, in addition to beginning stages.

Finally, the LUH argues for a new kind of experimental data in L2 research on the acquisition of aspect. Research on the brain signatures of aspect in (typically and atypically developing) adults, children, and even infants has used electrophysiological data (event related potentials [ERP]) and functional magnetic resonance imaging (fMRI) for at least 12 years (e.g., Baggio et al. 2008, Romagno et al. 2012). It is perhaps understandable that Slabakova (2006) could find no ERP study on the acquisition of L2 aspect (although L2 studies using ERP had been published since 1996). It is less understandable that the situation has not changed much since then. To my knowledge, electrophysiological and neuroimaging studies on L2 aspect using ERP and fMRI are still missing, yet they would be greeted by many as a much-needed change of perspective in the field. ERP data would be particularly revealing about L2 learners' developing aspectual representations. They would tell us not only what learners can do, but possibly also what they really know about what they are doing. This would mark the point when L2 research on the acquisition of aspect will really take a step forward: from assuming that an L2 predicate is telic or atelic to discovering whether and when the learner comes to know it.

REFERENCES

Andersen, Roger. 1984. *Second language: A cross-linguistic perspective*, Rowley: Newbury House.

Andersen, Roger. 2002. The dimensions of "Pastness". In *The L2 Acquisition of tense-aspect morphology,* ed. Rafael Salaberry and Yasuhiro Shirai, 79–106. Amsterdam: John Benjamins.

Andersen, Roger, and Yasuhiro Shirai. 1994. Discourse motivations for some cognitive operating principles. *Studies in Second Language Acquisition* 16(2): 133–156.

Andersen, Roger, and Yasuhiro Shirai. 1996. The primacy of aspect in first and second language acquisition: The pidgin/creole connection. In *Handbook of second language acquisition,* ed. William C. Ritchie and Tey K. Bhatia, 527–570. New York: Academic Press.

Ayoun, Dalia, and Jason Rothman. 2013. Generative approaches to the L2 acquisition of temporal-aspectual-mood systems. In *The research design and methodology in studies on second language tense and aspect,* ed. Rafael Salaberry and Llorenc Comajoan, 119–156. Boston: De Gruyter.

Baggio, Giosuè, Michiel van Lambalgen, and Peter Hagoort. 2008. Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language,* 59(1): 36–53.

Bardovi-Harlig, Kathleen. 1995. A narrative perspective on the development of the tense/aspect system in second language acquisition. *Studies in Second Language Acquisition,* 17(2): 263–291.

Bardovi-Harlig, Kathleen. 2000. *Tense and aspect in second language acquisition: Form, meaning and use*. London: Blackwell.

Bardovi-Harlig, Kathleen, and Llorenç Comajoan-Colomé 2020. The aspect hypothesis and the acquisition of L2 past morphology in the last 20 years. A state-of-the-scholarship review. *Studies in Second Language Acquisition* 42(5): 1137–1167. doi:10.1017/S0272263120000194

Barki, Pazit, Sandra Gorelli, Sabrina Machetti, Maria Pia Sergiacomo, and Beatrice Strambi. 2003. *Valutare e certificare l'italiano di stranieri: I livelli iniziali* [Evaluation and assessment of L2 Italian: The beginner levels]. Perugia: Guerra.

Borer, Hagit. 2004. The grammar machine. In *The unaccusativity puzzle,* ed. Artemis Alexiadou, Elena Anagnostopoulou, and Martin Everaert, 288–331. Oxford: Oxford University Press.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Cruse, David. 1997. *Lexical semantics*. Oxford: Oxford University Press.

Dik, Simon. 1994. Verbal semantics in functional grammar. In *Tense, aspect and action: Empirical and theoretical contributions to language typology,* ed. Carl Bache, Hans Basbøll, and Carl Erik Lindberg, 23–42. Berlin: Mouton de Gruyter.

Domínguez, Lucia, Nicole Tracy-Ventura, Maria Arche, Rosamund Mitchell, and Florence Miles. 2013. The role of dynamic contrasts in the L2 acquisition of Spanish past tense morphology. *Bilingualism: Language and Cognition,* 16(3): 558–577.

Dowty, David. 1979. *Word meaning and Montague Grammar*. Dordrecht: Reidel.

Dowty, David. 1986. The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? *Linguistics and Philosophy*, 9(1): 37–61.

Filip, Hana. 2004. The telicity parameter revisited. In *Semantics and Linguistic Theory* (*SALT XIV*), ed. Robert Young, 92–109. Ithaca: CLC Publications. Department of Linguistics, Cornell University.

Filip, Hana. 2012. Lexical aspect. In *The Oxford handbook of tense and aspect,* ed. Robert Binnik, 721–751. Oxford: Oxford University Press.

Folli, Raffaella, and Gillian Ramchand. 2005. Prepositions and results in Italian and English: An analysis from event decomposition. In *Perspectives on Aspect,* ed. Henk Verkuyl, Henriette de Swart, and Angeliek van Hout, 81–106. Dordrecht: Springer.

Forster, Kenneth. 2000. The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition,* 28(7): 1109–1115.

Giacalone Ramat, Anna, ed.. 2003. *Verso l'italiano*[Towards Italian]. Roma: Carocci.

Giacalone Ramat, Anna, and Stefano Rastelli. 2013. The qualitative analysis of actionality in learner language. In *Research design and methodology in studies on second language tense and aspect*, ed. Rafael Salaberry and Llorenc Comajoan, 391–422, Berlin: Walter de Gruyter.

van Hout, Angeliek. 2008. Acquiring telicity crosslinguistically: On the acquisition of telicity entailments associated with transitivity. In *Crosslinguistic perspectives on argument structure,* ed. Melissa Bowermann and Penelope Brown, 255–278. New York: Lawrence Erlbaum Associates.

Jegerski, Jill, and Bill VanPatten. 2013. *Research methods in second language psycholinguistics*. New York: Routledge.

Krifka, Manfred. 1992. Thematic relations as links between nominal reference and temporal constitution. In *Lexical matters,* ed. Ivan Sag and Anna Szabolsci, 29–53. Stanford: CSLI Lecture Notes.

Krifka, Manfred. 1998. The origins of telicity. In *Events and grammar*, ed. Susanne Rothstein, 197–235. Dordrecht*:* Kluwer.

Lardiere, Donna. 2003. The comparative fallacy revisited: A reply to Lakshmanan and Selinker (2001). *Second Language Research* 19(2): 129–143.

Li, Ping. 2000. The acquisition of lexical and grammatical aspect in a self-organizing feature-map model. In *Proceedings of the 22 Annual conference of the Cognitive Science Society,* ed. Leila Gleitman and Joshi Aravind, 304–309. Hillsdale: Erlbaum.

Li, Ping. 2002. Emergent semantic structure and language acquisition: A dynamic perspective. In *Cognitive neuroscience studies of the Chinese language,* ed. Che-Kan Leong, Gao Ding-Guo, and Henry Kao, 79–98. Hong Kong: The Hong Kong University Press.

Li, Charles, and Sandra Thompson. 1981. *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

Mackey, Alison, and Susan Gass. 2005. *Second language research. Methodology and design.* New York: Routledge.

McManus, Kevin. 2013. Prototypical influence in second language acquisition: What now for the Aspect Hypothesis. *International Review of Applied Linguistics and Language Teaching* (*IRAL*) 51(3): 299–322.

Montrul, Silvina. 2002. Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition* 5(1): 39–68.

Mourelatos, Alexander. 1978. Events, processes, and states. *Linguistics and Philosophy* 2(3): 415–434. doi:10.1007/BF00149015

Nishi, Yukio. 2012. A temporal approach to motion verbs. 'Come' and 'go' in English and East Asia Languages. In *Space and time in languages and cultures. Linguistic Diversity*, ed. Luna Filipovic and Kasia Jaszczolt, 395–413, Amsterdam: John Benjamins.

Nishi, Yukio. 2016. What learners know about lexical aspect in L2: Motion verbs *kuru* 'come' and *iku* 'go' and the acquisition of imperfective *-teiru* in Japanese. In *Cognitive-functional approaches to the study of Japanese as a second language*, ed. Kaori Kabata and Kiyoto Torotani, 57–87, Berlin: De Gruyter Mouton.

Nishi, Yukio, and Yasuhiro Shirai. 2019. Verb learning and the acquisition of aspect. L1 transfer of verb semantics. *Linguistic Approaches to Bilingualism* 11(3–4): 323–367. <https://doi.org/10.1075/lab.18084.nis>

Plonsky, Lule, Emma Marsden, Dustin Crowther, Susan M. Gass, and Patti Spinner. 2019. A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research* 36(4): 583–621. doi: 10.1177/0267658319828413

R Core Team. 2015. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Available at <http://www.R-project.org/>

Rastelli, Stefano. 2008. A compositional account of L2 verb actionality and the Aspect Hypothesis, *Lingue e Linguaggio* VII: 261–289.

Rastelli, Stefano. 2009. Lexical aspect too is learned: Data from Italian learner corpora. In *Multilingualism*, ed. Anna Saxena and Ake Viberg, 272–282, Uppsala: Edita Astra Varos.

Rastelli, Stefano. 2019. The imperfective paradox in a second language: A dynamic completion-entailment test. *Lingua* 231, doi.org/10.1016/j.lingua.2019.06.010

Rastelli, Stefano. 2020a. Contingency learning and perfective morpheme productivity in L2 Italian: A study on lexeme–morpheme associations with ΔP, *Corpus Linguistics and Linguistic Theory* 16(3): 459–486. <https://doi.org/10.1515/cllt-2019-007>

Rastelli, Stefano. 2020b. *The acquisition of aspect in a second language.* Cambridge: Cambridge University Press.

Rastelli, Stefano. 2021. Initial L2 learners may ignore or disregard whether predicates are telic or atelic: Counterevidence to the Aspect Hypothesis, *Lingue e Linguaggio*, XX(1); 81–110

Rastelli, Stefano, and Mirta Vernice. 2013. Developing Actional competence and the building blocks of telicity in L2 Italian. *International review of applied linguistics in language teaching* (IRAL) 51: 55–75.

Romagno, Domenica, Giuseppina Rota, Emiliano Ricciardi, and Pietro Pietrini. 2012. Where the brain appreciates the final state of an event: The neural correlates of telicity. *Brain and Language*, 123(1): 68–74 <http://dx.doi.org/10.1016/j.bandl.2012.06.003>

Salaberry, Rafael. 2008. *Marking past tense in second language acquisition: A theoretical model.* New York: Continuum.

Salaberry, Rafael. 2011. Assessing the effect of lexical aspect and grounding on the acquisition of L2 Spanish past tense morphology among L1 English speakers. *Bilingualism: Language and Cognition*, 14 (2): 184–202.

Salaberry, Rafael., and Llorenç Comajoan. 2013. *Research design and methodology in studies on L2 Tense and Aspect*. Boston: De Gruyter Mouton.

Salaberry, Rafael, and Custodio Martins. 2014. Crosslinguistic transfer of core aspectual conceptualizations in Portuguese and Spanish: Theoretical and methodological factors. In *Portuguese-Spanish interfaces: Diachrony, synchrony and contact*, ed. Patricia Amaral and Ana Maria Carvalho, 335–355, Amsterdam: John Benjamins.

Samu, Borbala. 2020. La grammatica cognitiva e l'insegnamento del tempo e dell'aspetto in italiano L2[Cognitive Grammar and the teaching of Tense and Aspect in L2 Italian]. *Italiano LinguaDue,* 12(1): 209–236.

Di Sciullo, Anna Maria, and Roumyana Slabakova. 2005. Quantification and aspect. In *Perspectives on Aspect,* ed. Henk Verkuyl, Henriette De Swart, and Angeliek van Hout, 61–80. Dordrecht: Springer.

Shirai, Yasuhiro. 2013. Defining and coding data: Lexical aspect in L2 studies. In *Research design and methodology in studies on L2 tense and aspect,* ed. Rafael Salaberry and Llorenc Comajoan, 271–308. Boston: Walter De Gruyter.

Slabakova, Roumyana. 2001. *Telicity in the second language*. Amsterdam: John Benjamins.

Slabakova, Roumyana. 2006. Is there a critical period for semantics? *Second Language Research*, 22(3): 302–338.

Slobin, Dan. 1979. *Psycholinguistics*. Glenview: Scott, Foresman and Company.

De Swart, Henriette. 1998. Aspect shift and coercion. *Natural Language and Linguistic Theory* 16(2): 347–385.

Talmy, Leonard. 1978. The relation of grammar to cognition: A synopsis. *Proceedings of the 1978 workshop on Theoretical issues in natural language processing* (TINLAP '78):14–24 <https://doi.org/10.3115/980262.980266>

Tatevosov, Sergey. 2002. The parameter of actionality. *Linguistic Typology* 6(3): 317–401.

Tong, Xi, and Yasuhiro Shirai. 2016. L2 acquisition of Mandarin zai and –le. *Chinese as Second Language Acquisition Research* (*CASLAR* 5): 1–26.

van Valin, Robert. 2005. *Exploring the syntax semantics interface*. Cambridge: Cambridge University Press.

Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca: Cornell University Press.

Verkuyl, Hank. 1993. *A theory of aspectuality*. Cambridge: Cambridge University Press.

Viberg, Åke. 2002. Basic verbs in lexical progression and regression. In *An integrated view of language development: Papers in honor of Henning Wode,* ed. Petra Burmeister, Thorsten Piske, and Andreas Rohde, 109–134. Wissenschaftlicher: Verlag Trier.

Wulff, Steffi, Nick Ellis, Kathleen Bardovi-Harlig, Chelsea Leblanc, and Ute Römer. 2009. The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal,* 3(3): 354–369. doi:10.1111/j.1540-4781.2009.00895.x

CORPORA

CLIP (Corpora e Lessico di Italiano Parlato)
Corpus Pavia
ItTenTen
LIP (Lessico dell'Italiano Parlato)