Stefano Rastelli*

Contingency learning and perfective morpheme productivity in L2 Italian: A study on lexeme–morpheme associations with ΔP

<https://doi.org/10.1515/cllt-2019-0071>

Abstract: This study utilized unidirectional association score ΔP to track perfective morpheme productivity in longitudinal spoken L2 Italian data. Research questions concerned whether early L2 perfectives were contingent upon telicity of predicates, whether lexeme–morpheme association changed as proficiency increased, and whether distribution of perfectives in the L1 input affected the patterns of morpheme emergence. Results showed that (i) the productive use of the perfective was contingent upon a few, infrequent telic predicates but also upon some actionally underspecified, very frequent general-purpose ones; (ii) a generalized decrease in association scores over time accompanied the productivity of the perfective morpheme; (iii) distribution of perfectives in L2 data did not reflect distribution in the L1 input. The statistical analysis adopted in this study is replicable to other domains where contingency of stem-affix alternations may provide cues for observing the developing L2 grammar

Keywords: contingency learning, learner corpora, lexeme-construction association, lexical aspect hypothesis

1 Introduction: Topic and motivation

Much corpus-linguistic work on second language acquisition (SLA) underutilizes the probabilistic information corpora may offer, “limiting itself to comparing absolute and relative frequencies of types and tokens” (Gries 2018a: 734). Frequency comparison alone is not informative enough. Items that are less frequent in absolute could be acquired earlier, more easily, and more stably than more frequent items. Other than raw frequency, other measures can inform us on how distribution in the input impacts L2 acquisition: dispersion, predictability and surprisal, recency, salience and association/contingency (e.g. Gries 2015; Gries and Ellis 2015; Ellis 2016). This study utilizes the unidirectional

*Corresponding author: Stefano Rastelli, Department of Humanities, Università di Pavia, Pavia, Italy, E-mail: stefano.rastelli@unipv.it

measure of association ΔP to explore contingency learning (CL) of the perfective morpheme in Italian learner data. There are two motivations for this study. First, CL is important in the current debate on L1 and L2 acquisition. According to usage-based approaches, grammatical constructions too “have a meaning” because speakers tend to associate them more frequently and stably with some lexical items – e.g. specific classes of verbs – rather than with others (e.g. Stefanowitsch forthcoming: 268). Learners’ sensitivity to lexeme-construction contingency could trigger the acquisition of grammatical categories. On the contrary, generative approaches maintain that formal (ϕ) features – such as the perfective – are learned independently of the lexicon. Some generative linguists accept the idea that the process of acquisition of ϕ -features can be permeable to frequency-related factors (e.g. Yang 2018), but not to the extent that the acquisition of such features is contingent on the meaning of the items in which they are instantiated.

The second reason of interest is that the fluctuations of contingency-based association scores over time can provide a new method for assessing morpheme productivity in SLA. This study examines whether and how morpheme–lexeme contingency changes over time in L2 learners’ developmental path. The weakening of contingency-based unidirectional association scores may indicate that learners have become capable of abstracting the construction away from specific lexical entries and of generalizing the grammatical pattern underneath.

2 Background: Contingency learning

CL is a probability-based mechanism for learning whether relations between events are casual or noncasual (Beckers, De Houwer and Matute 2007: 289; Shanks 2007). Subjects tend to label co-occurring events as noncasual when the probability P of getting an outcome O (e.g. thunder) given a cue C (e.g. lightning) ($P(O|C)$) is very high. Noncasual cue–outcome relationships, once identified, trigger category formation (Gluck and Bower 1988). Categories are formed when subjects first group co-occurring events (e.g. words) basing on adjacency and resemblance and then assign these events category membership based on noncasual relationships (Reeder et al. 2010; Ramscar et al. 2010). In first and SLA, CL might sustain the process of category formation by exploiting the frequent co-occurrence between a lexeme and a grammatical construction (Goldberg 2006). The current study investigates whether L2 learners’ sensitivity to frequency of lexeme–construction associations may trigger the acquisition of aspectual categories, such as the perfective.

2.1 Contingency learning and the acquisition of L2 aspect

CL occurs when L2 learners realize that a semantic feature of the verb lexeme and the perfective construction are a close match. The contingent notions are those of telicity – a feature of lexical aspect¹ (LA) – and of perfectivity, a feature of grammatical aspect (GA). LA refers to how the events designed by the predicates flow in time. Speakers may represent such events as having an endpoint (e.g. “arrive home”) or as flowing homogeneously (“walk a mile”), as being instantaneous (“explode”) or prolonged (“rain”), iterative (“cough”) or punctual (“sneeze”), as being characterized by gradual (“rise”) or sudden change (“the light goes down”). The best-known classification of LA follows the Vendler–Mourelatos hierarchy (Vendler 1967; Mourelatos 1978). The hierarchy divides predicates into four categories (states, activities, accomplishments, and achievements) depending on the different clustering of the semantic features [± static], [± durative], and [± telic] in the meaning of those predicates. Italian presents no overt marking of LA features; therefore, such features are specified lexically rather than morphologically. CL in this paper concerns specifically the feature of telicity. All telic predicates have a culmination point in their semantic representations. Both interval semantics (Dowty 1979) and the mereological approaches (Champollion and Krifka 2016) – although from very different perspectives – differentiate between telic and atelic events just by the presence or absence of a culmination point. Krifka (1992) coupled a culmination point’s presence with non-homogeneous events and its absence with homogeneous ones. The event of “pushing a chair” is homogeneous because it lacks any qualitatively different sub-interval throughout its duration. By contrast, “painting a wall” is not homogeneous because it involves a special sub-interval. This is the moment when the wall is completely painted. Thus, telic events have a richer, differentiated (non-homogeneous) internal structure because it includes a result.

GA refers to the speaker’s perspective on the event, that is, it reflects the speaker’s decision on how an event can be described. An Italian speaker can use the perfective to depict an event as a closed interval, viewed from beginning to end. Alternatively, a speaker can use the imperfective to depict an event as an open interval whose conclusion is out of sight (Comrie 1976, p. 16; Klein 1994; Smith 1997). GA in Italian is partially overt. In the past tenses, the [±perfective] distinction is encoded in morphology (Bertinetto 2001). Having evolved from an original present perfect value and having absorbed the aoristic (punctual) value

¹ LA is also called actionality, inherent aspect, semantic aspect, inner aspect, situation aspect, predication aspect, eventuality type, state of affairs, and Aktionsart.

of a competing past tense (the *passato remoto*), the modern Italian *passato prossimo* (e.g. *Mario ha giocato a tennis* “Mario played tennis”) is a compound tense which encodes the perfective past. This form may also express the meaning of a present perfect in English. By contrast, the Italian *imperfetto* (e.g. *Mario giocava a tennis* “Mario used to play/was playing/played tennis”) encodes the imperfective value and conveys unbounded, habitual, or ongoing past events.

2.2 CL and the emergence of the perfective

Two conditions seem necessary for CL to occur. The first is distinctiveness. The quality that makes a given lexeme a perfect match for the perfective can only be shared by a subset of predicates and not by all. The smaller this subset is, the faster the overarching category will be acquired (Nosofsky 2014). Distinctiveness boosts category formation in general human cognition, not only in SLA. An oft-cited example is that subjects learning the category /birds/ would initially focus on birds’ distinctive features (e.g. wings) rather than on what is shared among birds and other animals (e.g. eyes) (Tversky and Gati 1978; Wulff et al. 2009: 356).

The second condition that regulates CL is its transiency. The importance of contingent associations diminishes as experience gets richer and learning progresses. If the strength of association between a cue (wing) and the outcome (/bird/) is the only factor driving the learning process, category membership will be restricted to few prototypical, perfectly matching exemplars. Generalizations at later stages of learning will break prior association by introducing new, less prototypical members (e.g. not all animals with wings are birds). In this paper, transiency of CL – a decrease of association scores over time – is taken as a cue of increasing morpheme productivity and of developing L2 learner’s grammar. Morpheme productivity is an affix’s ability to generate new forms systematically. The more an affix participates in different lexemes, the more productive it is. In the literature, morpheme productivity is usually considered a direct function of the type frequency of the relevant constructions (e.g. Croft and Cruse 2004, p. 309; Gries and Ellis 2015, p. 234). In the case of the current study, the type frequency of the Italian perfective construction is the number of different verb lexemes with which the perfective past (the *passato prossimo*) co-occurs in a learner corpus. When the perfective construction’s type frequency is low, the perfective morpheme is unlikely to be represented in a L2 Italian learner’s competence. L2 learners are able to represent the perfective morpheme if they conceive of it as separate and independent from the various features of the specific lexemes in which the morpheme occurs. In order to distinguish this qualitative change in a learner’s L2 development, this paper analyzes lexeme–morpheme contingency in

longitudinal learner data. The consistent weakening of the unidirectional lexeme–morpheme association score (ΔP , Section 4.4) could quantify the extent to which learners are capable of abstracting the construction away from specific lexical entries and generalizing the grammatical patterns underneath. The analysis of contingency is more informative regarding morpheme productivity than the analysis of type frequency, for two reasons. First, unlike type frequency, contingency can tell us whether the relationship between a lexeme and a morpheme is different from all relationships among lexemes and morphemes in a corpus. Second, it can reveal also the directionality of association: whether the morpheme attracts the lexeme or the other way around.

2.3 Related and unrelated CL and the acquisition of the perfective

CL comes in two flavors: related contingency learning (RCL) and unrelated contingency learning (UCL). In RCL, the cue (in our case, telicity) and outcome (in our case, the perfective morpheme) have one or more traits in common (e.g. the idea of event delimitation, see Section 2.1). On the contrary, in UCL, the cue and outcome do not share any traits. The cognitive principles governing RCL and UCL are *common coding* and *separate coding*, respectively (Hommel and Prinz 1996; Hommel et al. 2001). Common coding reflects the possible existence of privileged linkages between a cue and its more compatible or more frequent responses. Separate coding reflects the view that stimuli and responses “are entirely different and incommensurate things that have nothing to do with each other” (Prinz 2018: 147). Learning based on separate coding can in principle link anything to anything. Unrelated items “offer themselves as ideal candidates for implementing mapping rules as sets of arbitrary linkages between stimulus and response codes” (Prinz 2018: 147). Common coding (underlying RCL) and separate coding (underlying UCL) inspire different explanations of how L2 Italian learners may acquire the perfective. A developmental theory assuming RCL and common coding of lexeme–morpheme association is the Lexical Aspect Hypothesis (LAH).

2.4 Related contingency: The lexical aspect hypothesis (LAH)

The LAH proposes that L2 learners acquire tense and aspect morphology of the target-language (TL) asymmetrically, with some predicates learned earlier than others (Andersen and Shirai 1994, 1996; Salaberry and Comajoan 2013). The

acquisition of the perfective–imperfective distinction would be modulated by learners’ knowledge of LA of L2 predicates. The LAH predicts that learners will associate the perfective morpheme first with telic and later with atelic predicates. As Andersen (2002: 78) put 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 would hold for the imperfective morpheme. While LA constrains the early emergence of perfective and imperfective verb morphology, at later stages of L2 acquisition, verb morphology would spread regardless. Different explanations have been proposed to account for the preferred patterns of association between telic predicates and perfective morphology. One explanation refers to the cognitive principle of prototypicality and the presence of a distributional bias in the input reflecting this principle (Andersen and Shirai 1994: 133). 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 they are congruent concepts. The inherent culmination point that characterizes telic predicates (such as the Italian verb *cadere* “fall”) is more compatible with bounded or “terminative” events presented at the past perfective (e.g. the Italian Passato Prossimo *è caduta* “(she) has fallen/(she) fell”) than with unbounded or “non-terminative” events presented at the past imperfective (e.g. the Italian Imperfetto *cadeva* “(she) fell/was falling”). Moreover, prototypical associations are also more frequent in the TL input than non-prototypical ones (Li and Shirai 2000). This skewed distribution would reinforce the effectiveness of the prototypicality principle in language acquisition.

2.5 Unrelated contingency: Lexical underspecification and generality of meaning

Under the perspective of UCL, telicity of predicates does not count for the emergence of the perfective morpheme. Instead, for a lexeme to attract the perfective morpheme – and the other way round – two factors are important: (a) the generality of predicate’s meaning and (b) its actional underspecification. As to the first factor, L2 learners are expected to attach early perfective morphemes preferably to “umbrella-verbs”: general verbs whose meanings include those of similar, more specific verbs. General verbs are very frequent because the more general the verb’s meaning, the more likely it will be used in various contexts. Research has found that L2 learners over-use some general verbs in place of verbs within the same

semantic category with more specific meanings or different LA categories. Viberg (2002) observed that in learner data, basic and nuclear verbs (such as “go”, “come”, “see”, “say”) tend to replace more specific verbs, especially in early stages of acquisition. For example, in the corpus Pavia, especially initial L2 learners very often over-extend the basic, verb *andare* “go” to more specific motion verbs such as *arrivare* “arrive”, *venire* “come”, *volare* “fly”, *entrare* “enter”, *uscire* “exit”, *salire* “go up”, *scendere* “go down” (Rastelli 2008).

The second expectation under the perspective of UCL is that early L2 learners will attach perfective morphemes especially to actionally² underspecified verbs. Actional underspecification characterizes predicates that can enter most or even all four Vendlerian LA categories. Basic verbs – like those seen above – are typically also lexically underspecified, because they can be telic or atelic depending on the context. In the Corpus Pavia, the most frequent predicates (such as *fare* “make, do”, *dire* “say”, *dare* “give”, *andare* “go”, and *prendere* “take, get”) can be telic or atelic depending on object NP and adjuncts. For example, *fare fatica* “struggle” is atelic whereas *fare fuoco* “shoot” is telic; *dare una festa* “party” is atelic whereas *dare un bacio* “kiss” is telic. In child language acquisition, evidence suggests that high-frequency, underspecified verbs are learned early. Theakston et al. (2004) listed the following light verbs that would be acquired first in English: bring, come, do, get, give, go, make, put, and take. The prediction based on UCL is that early perfectives in the corpus Pavia will be general-meaning, basic and actionally underspecified because all verbs belonging to these categories can fit different contexts and suffice well for different communicative tasks.

2.6 The role of L1 input

The emergence of the perfective may be contingent not only upon telicity or actional underspecification, but also upon the frequency of perfective predicates in L1 input. Wulff et al. (2009) focused on two native corpora: the spoken section of the British National Corpus (BNC, 10 million words) and the Michigan Corpus of Academic English (MICASE, 1.7 million words). Using multiple distinctive collxeme analysis (Gries and Stefanowitsch 2004) and ΔP , they found that the verbs first learned in the perfective past tense were highly telic and frequent in and distinctly associated with the perfective past tense in the L1 input. Tracy-Ventura and Cuesta Medina (2018) found an identical distributional bias in L1

2 The adverb “actionally” refers to the term “Actionality”, which some authors utilize to denote LA as distinct from the notion of “Aspect” in Slavic languages (e.g. Bertinetto 2001).

Spanish, with several telic predicates more often occurring in the preterit and several atelic predicates in the imperfect. The role of L1 frequency could be modulated by other factors. For example, early perfective morphemes could not be attached just to the most frequent verbs in the L1 input, but also to verbs that are more relevant to the contexts of elicitation and to the learner. Such verbs likely are the most frequent in the learner corpus, not necessarily in the L1 input because they could reflect the peculiar elicitation context.

3 Research questions

In the previous sections, we focused on three factors that may affect the emergence of the perfective morpheme in longitudinal learner data: LA (telicity as opposed to generality of meaning and to actional underspecification), L1 frequency and transiency. These factors motivated three research questions (RQ) in this study. RQ₁ contrasts the predictions for the emergence of the perfective morpheme based on RCL and UCL. Early perfectives could be contingent upon telicity of verb lexemes or upon actional underspecification and generality of meaning (see Sections 2.4 and 2.5). RQ₂ investigates the role of L1 input. Early perfectives could be those more frequent in the L1 input, or the most frequent in the corpus (e.g. the most relevant for the communicative contexts and for the elicitation tasks) (see Section 2.6). RQ₃ concerns whether the association measure ΔP in longitudinal data change over time, given that transiency is a peculiar feature of CL (Section 2.3):

- RQ₁: Is the emergence of the perfective morpheme in L2 data contingent upon related factors (telicity) or upon unrelated factors (generality of meaning and actional underspecification)?
- RQ₂: Does the distribution of perfectives in the L1 input play a role in the emergence of the perfective in L2 data?
- RQ₃: Do lexeme–morpheme contingency scores change over time in longitudinal learner data?

4 Method

4.1 Target feature

The current paper concerns the emergence of the perfective in L2 Italian. In this study's dataset, all perfective morphemes occur in the Italian perfective past (the

passato prossimo), whereas none occurs in the aoristic-perfective (the *passato remoto*). The Italian *passato prossimo* comprises an auxiliary (the inflected form of *avere* “have” or *essere* “be”) and a past participle that hosts the perfective morpheme.

4.2 L2 data

We explored CL in the *Corpus Pavia*, the largest and best known longitudinal learner Italian corpus to date (~700,000 tokens, ~15,000 types overall³) (Giacalone Ramat 2003). It contains transcriptions of about 120 hours (2 to 10 hours for each learner) of oral interviews of 22 Italian L2 learners from 11 different L1 backgrounds spanning five typological families. During the interviews, learners engaged in spontaneous and semi-structured conversations and tasks with Italian interviewers. Conversation topics varied a lot across both learners and interviews and included everyday life, cultural differences, countries of origin, leisure activities, interpersonal relations, and features of Italian. Supervised elicitation tasks were also used at times and included description of pictures and oral retelling of picture-stories and video excerpts from the film *Modern Times*. Online Appendix 1 describes the Corpus Pavia’s composition in terms of learner- and data-related dimensions. In Online Appendix 2, the procedure for formatting and coding data is described. After manual normalization-tokenization and after cross-checked lemmatization and tagging with Sketch Engine,⁴ each learner’s file comprised between 124 and 2,887 finite verb forms (983 at average), totaling 22,109 verb tokens and 5,540 perfective tokens stemming from 304 perfective types. For the present study, we selected a sample of 39 perfectives having frequency ≥ 25 in the corpus. The cutoff point was chosen after visual inspection of the Zipfian curve with the purpose of guaranteeing the sample’s homogeneity, manageability, representativeness, and the density of data. After that threshold, the absolute frequencies of perfectives predicates drop considerably, and the Zipfian curve approaches its inflection point. The resulting sample of 39 perfective predicates contained 3,940 perfective tokens (see online Appendix 5).

³ Excluding interviewers’ prompts and contributions, repetitions, and false starts.

⁴ <https://auth.sketchengine.eu>.

4.3 Data aggregation

The purpose of this study was to identify changing patterns of lexeme-perfective associations in longitudinal learner data. Data sparseness – the fact that perfective types and tokens were unevenly distributed across both interviews and learners – might have undermined this goal. For example, although the telic perfective *venuto* “come” has very high absolute frequency in the corpus Pavia, some learners used it only at early interviews, others used it only at late interviews and some did not use it at all. Having many, few or zero-occurrence of *venuto* at early interviews might not depend on the main independent variable in our study (learner’s sensitivity to related or unrelated CL) but on the kind of task and on the topic of interviews, whose distribution did not follow a predictable pattern across interviews. To minimize this contextual bias, longitudinal data in this study were aggregated. Period of interview was chosen as the most comprehensive and neutral criterion for aggregating data across learners. For each learner, interviews were grouped into three periods: “early”, “mid”, and “late”. Each period was balanced within-learner for number of interviews (from 1 to 6). Although the overall number of verb tokens changed across periods, the time-span (ranging from 1 week to 4 months) in within-learner interviews was kept constant. The choice of aggregating data from different learners across three equally distant periods is methodologically questionable. This seems the kind of sampling that Gries and Stoll (2009) have argued against for acquisition data by proposing a Variability-based Neighbor Clustering (VNC) approach and that Stoll and Gries (2009) have argued against by proposing regression-based approaches that determine the temporal stages from the behavior of the regression models. In fact, the cut-offs on the temporal continuum (in our case, “early”, “mid” and “late” interviews) are arbitrary and that the existence of developmental stages is assumed *a priori* instead of emerging “bottom-up” from the data (Gries and Stoll 2009: 223). However, there are cases when arbitrariness does not stem from theoretical preconceptions (as it is claimed by Gries and Stoll 2009: 219) but it is a necessary expedient in order to minimize the impact of intervening variables that cannot be controlled for. In order to elicit spoken productions in the corpus Pavia, researchers utilized an open repertoire of tasks and topics. The choice of whether or not – and of when – to utilize a pre-determined topic/task was left to interviewers’ decision. For example, some learners were asked to describe how they arrived in Italy or to tell a scene of *Modern Times* – two pre-determined topics that, respectively, likely elicit or dissuade the use of *venire* “come” at the perfective past – indifferently at early, mid or late interviews. Other learners were instead asked just to talk about their plans for the future. Since topics and tasks were either intermittent or interspersed randomly across interviews, a

clustering algorithm that agglomerate temporal adjacent values of the perfective predicates basing on a similarity metrics – like the VNC – would risk to exchange the results of a biased elicitation procedure for the presence or absence of cues of a learner’s developing competence over time.⁵ Sampling data from large intervals – rather than from adjacent points – and aggregating different within- and between-learners interviews minimize the impact of the heterogeneity of the elicitation tasks and topics. The logic underlying our choice is simple: the larger the sampling intervals, the more topics and tasks they include and the least their differences will impact the outcomes. If the analysis of ΔP can actually surface patterns of lexeme-morpheme contingency over time, these will have emerged despite of – and not because of – the heterogeneity of tasks and topics. While bearing in mind the risks that may come with this methodological choice, we considered that in this case data aggregation is not a “bug”, but rather a “feature” of the current study and that it represents the lesser evil.

4.4 The unidirectional association measure ΔP

To identify the cues of CL in learners’ production over time, we utilized a unidirectional, proportion-based association scale ΔP (Ellis 2006; Ellis and Ferreira-Junior 2009: 198). Unlike bidirectional association measures, ΔP can separately assess each item’s contribution to the overall strength of association by comparing two kinds of transition probabilities between two items, in our case, the lexeme (a verb) and the construction (the perfective). One measure (“reliance”) compares the relative frequency of the construction with the lexeme to the relative frequency of the construction without the lexeme. The other measure (“attraction”) compares the relative frequency of the lexeme with the construction to the relative frequency of the lexeme without the construction. The starting point for the calculation of ΔP is a contingency table like Table 1, where values *a* through *d* correspond – in our study – to the co-occurrence frequencies between a verbal lexeme (*x*) and the perfective morpheme (*y*) in the corpus Pavia:

Cell (a) in Table 1 corresponds to the number of responses (the perfective) given the cue (e.g. a telic, or atelic, or underspecified verb lexeme). Cell (b) corresponds to

⁵ Clustering algorithms that – unlike VNC – do not keep into account adjacency would perform even worse. They would aggregate very distant values – e.g. those sharing topics and tasks – luring researchers to conclude (maybe wrongly) that time of interview did not affect the number of perfectives.

Table 1: A contingency table.

	Presence of <i>y</i> (response)	Absence of <i>y</i> (no response)
Presence of <i>x</i> (cue)	a	b
Absence of <i>x</i> (no cue)	c	d

all cues (e.g. instances of the lexeme without the response (the perfective). Cell (c) corresponds to the number of responses without the cue (e.g. all perfectives in the interviews except (a)); and cell (d) corresponds to all predicates uttered by the learner in the whole corpus not including all perfectives and all forms of the lexeme. The formula for calculating ΔP has two halves. The first half of the formula⁶ treats the lexeme as the cue and the perfective as the response. It shows the difference between the frequency of the perfective with and without the lexeme and reflects “reliance”. The second⁷ treats the perfective morpheme as the cue and the lexeme as the response; it shows the difference between the frequency of the lexeme with and without the perfective. This value reflects attraction. ΔP is a scale (based on proportions), not a test of significance, so there is no minimal threshold value (Ellis 2012: 28). The value of ΔP in the current study lies just in the comparative information it provides (see discussion).

4.5 Coding the lexical aspect of L2 predicates

To test whether the emergence of the perfective morpheme is contingent upon telicity rather than actional underspecification, we coded the LA of L2 predicates. Unlike Wulff and colleagues (2009), we did not code LA by looking at semantics alone (e.g. “the sense of the verb that comes to mind first”, Wulff et al. 2009: 356) for two reasons. First, LA is often compositional (Verkuyl 1993): it results from the interplay between the inflected verb, its arguments, and its collocates, outside the VP and up to the sentence level. Second, looking at L2 verbs alone augments the risk of committing the comparative fallacy, imposing the target-language perspective on a learner’s developing competence (Bley-Vroman 1983; Lardiere 2003). If the (idealized) meaning of the predicate in isolation is the only element used for coding LA, the risk of learner and native speaker (NS) judgments diverging increases. The procedure for coding data and

⁶ Reliance = ΔP cuelex: $[a/(a + b)] - [c/(c + d)]$.

⁷ Attraction = ΔP cuecx: $[a/(a + c)] - [b/(b + d)]$.

the results of the procedure are described in online Appendix 3. The final labeling of VPs as telic, atelic or underspecified depended on the Cronbach’s α scores (displaying inter-rater agreement) that – in turn – correlated with the quantitative prevalence of actional interpretations for a given VP by the native Italian raters.

4.6 Coding general-purpose predicates

As we have seen (Sections 2.5 and 4.6) that the emergence of the perfective might not be contingent on telicity but on general-purpose predicates. To identify such predicates we selected from our sample eight perfectives that matched those quoted by Theakston et al. (2004) and Wulff et al. (2009). These are the Italian equivalents of the general-purpose English verbs *do/make, go, bring, put, see, take, come, and give*.⁸ Table 2 reports the list of general-purpose perfectives, their absolute frequency, their frequency rank and their LA, as it was coded by a panel of Italian NS (Section 4.5).

Table 2: Lexical aspect and frequency of general-purpose verbs in the Corpus Pavia.

Verbs	English translation	Token frequency	Rank	Lexical aspect category
<i>fatto</i>	Done	437	2	Underspec
<i>andato</i>	Gone	420	3	Underspec
<i>visto</i>	Seen	323	4	Underspec
<i>preso</i>	Taken	138	7	Underspec
<i>venuto</i>	Come	96	10	Telic
<i>portato</i>	Brought	61	17	Underspec
<i>dato</i>	Given	55	20	Underspec
<i>messo</i>	Put	35	35	Underspec

As Table 2 suggests, the NS sample rated most general-purpose predicates produced by learners as being actionally underspecified. These general-purpose predicates were also very frequent in the corpus. In many cases, the contextual relevance of such predicates for the elicitation task adopted in the corpus can explain their high frequency.

8 The verb *avuto* “had” was discarded because it has too many Italian equivalents.

5 Results

5.1 Contingency

RQ₁ asked if the emergence of the perfective morpheme in L2 data is contingent upon related factors (telicity) or upon unrelated factors (generality of meaning and actional underspecification). Tables 3 and 4 report the complete set of ΔP values at early interviews for the 39 sampled predicates, ordered, respectively, by reliance and by attraction.

In Figure 1, the ΔP values of lexeme-perfective attraction (x -axis) and reliance (y -axis) at early interviews are plotted. To make visual interpretation easier, we utilized the English translation of the Italian perfectives. Different colors represent different LA categories, as they were coded by Italian NSs.

Inspection of Figure 1 suggests that – in general – early L2 lexemes did not attract the perfective morpheme. In fact, most attraction values clustered around negligible values (0–0.025). Even the values of those few underspecified (*detto* “said”, *andato* “gone”, *visto* “seen”, *fatto* “done”) and telic (*arrivato* “arrived”, *capito* “understood”) lexemes that seemed to stand out were very low, around 0.015. Instead, the perfective morpheme relied (range = 0.4–0.9) on at least eleven lexemes, eight of which were telic and two actionally underspecified. There was a moderate to strong negative correlation between predicates frequency rank and attraction (Kendall $\tau = -0.52$), meaning that the most frequent perfectives in early interviews most attracted the perfective morpheme, even though – as observed above – attraction was very weak throughout. On the contrary, reliance and frequency ranking of early perfectives did not correlate. A Kruskal-Wallis test with a post-hoc Dunn’s test (*dunnTest* function in the FSA package of software R) showed that while reliance correlated with LA (the perfective morpheme relying more on telic than on underspecified or atelic lexemes, average between groups difference = 0.287, chi-squared = 15.539, $df = 2$, p value < 0.001), attraction was not determined by LA (chi-squared = 2.907, $df = 2$, $p = 0.23$).

Also the distribution of ΔP values of attraction and reliance across the sampled perfectives predicates was different. While the distribution of attraction values was Zipfian, the distribution of reliance values was not:

The information about Zipfian distribution can be relevant for the issue of distinctiveness of CL (Section 2.2). It would be expected that – in order to boost category formation – the perfective should be contingent only upon a small subset of highly reliable lexemes. On the contrary, Figure 2 shows that the reliance values in the corpus Pavia degrade gradually and are distributed quite evenly across a considerable number of predicates in the corpus. This suggests that – if attraction is

Table 3: List of 39 sampled perfectives ordered by “reliance”.

	Perfective	English translation	Reliance	Frequency rank in the corpus
1	<i>dimenticato</i>	Forgotten	0,828	16
2	<i>sbagliato</i>	Mistaken	0,743	27
3	<i>perso</i>	Lost	0,664	34
4	<i>finito</i>	Finished	0,645	10
5	<i>portato</i>	Taken	0,539	19
6	<i>morto</i>	Died	0,533	29
7	<i>dato</i>	Given	0,523	20
8	<i>incontrato</i>	Met	0,474	32
9	<i>arrivato</i>	Arrived	0,469	6
10	<i>capito</i>	Understood	0,458	5
11	<i>caduto</i>	Fallen	0,41	30
12	<i>tornato</i>	Returned	0,386	37
13	<i>lasciato</i>	Let	0,371	17
14	<i>imparato</i>	Learned	0,325	11
15	<i>passato</i>	Past	0,32	24
16	<i>trovato</i>	Found	0,308	8
17	<i>sposato</i>	Got married	0,295	33
18	<i>venuto</i>	Come	0,274	7
19	<i>uscito</i>	Exited	0,247	23
20	<i>fatto</i>	Done	0,239	1
21	<i>sentito</i>	Heard	0,234	15
22	<i>telefonato</i>	Phoned	0,219	39
23	<i>visto</i>	Seen	0,206	3
24	<i>detto</i>	Said	0,178	4
25	<i>scritto</i>	Written	0,175	22
26	<i>chiesto</i>	Asked	0,164	31
27	<i>cambiato</i>	Changed	0,145	35
28	<i>andato</i>	Gone	0,141	2
29	<i>comprato</i>	Bought	0,126	28
30	<i>parlato</i>	Talked	0,097	9
31	<i>lavorato</i>	Worked	0,071	13
32	<i>studiato</i>	Studied	0,059	21
33	<i>preso</i>	Taken	0,049	18
34	<i>dormito</i>	Slept	0,041	36
35	<i>pensato</i>	Thought	0,036	12
36	<i>mangiato</i>	Eaten	0,03	14
37	<i>piaciuto</i>	Liked	−0,005	26
38	<i>messo</i>	Put	−0,01	38
39	<i>voluto</i>	Wanted	−0,04	25

Table 4: List of 39 sampled perfectives ordered by “attraction”.

	Perfective	English translation	attraction	Frequency rank in the corpus
1	<i>fatto</i>	Do	0,101	1
2	<i>visto</i>	See	0,093	3
3	<i>andato</i>	Go	0,063	2
4	<i>detto</i>	Say	0,062	4
5	<i>capito</i>	Understand	0,057	5
6	<i>arrivato</i>	Arrive	0,05	6
7	<i>finito</i>	Finish	0,025	10
8	<i>venuto</i>	Come	0,025	7
9	<i>trovato</i>	Find	0,022	8
10	<i>imparato</i>	Learn	0,021	11
11	<i>dimenticato</i>	Forget	0,014	16
12	<i>portato</i>	Take	0,013	19
13	<i>parlato</i>	Talk	0,013	9
14	<i>dato</i>	Give	0,012	20
15	<i>lasciato</i>	Let	0,012	17
16	<i>sbagliato</i>	Mistake	0,011	27
17	<i>sentito</i>	Hear	0,011	15
18	<i>passato</i>	Pass	0,01	24
19	<i>morto</i>	Die	0,009	29
20	<i>caduto</i>	Fall	0,009	30
21	<i>uscito</i>	Exit	0,009	23
22	<i>incontrato</i>	Meet	0,008	32
23	<i>scritto</i>	Write	0,008	22
24	<i>lavorato</i>	Work	0,008	13
25	<i>perso</i>	Loose	0,006	34
26	<i>sposato</i>	Get married	0,006	33
27	<i>chiesto</i>	Ask	0,006	31
28	<i>comprato</i>	Buy	0,006	28
29	<i>studiato</i>	Study	0,005	21
30	<i>pensato</i>	Think	0,005	12
31	<i>tornato</i>	Return	0,004	37
32	<i>cambiato</i>	Change	0,004	35
33	<i>preso</i>	Take	0,004	18
34	<i>mangiato</i>	Eat	0,003	14
35	<i>telefonato</i>	Phone	0,002	39
36	<i>dormito</i>	Sleep	0,001	36
37	<i>piaciuto</i>	Like	0	26
38	<i>messo</i>	Put	0	38
39	<i>voluto</i>	Want	−0,007	25

Figure 1: Values of reliance (y) and attraction (x) at early interviews.

Figure 2: Zipfian curves of reliance (left) and attraction (right) at early interviews.

certainly distinctive – high reliance does not characterize a limited set of telic lexemes. To sum up, in order to answer whether the emergence of the perfective was contingent on LA, one should consider separately attraction, reliance and their relationship with frequency. Our results suggested a three-ways interaction. The lexemes that attracted most the perfective morpheme in early interviews were few, very frequent and lexically underspecified. However, those values were very low all across the board. Instead, early perfective morphemes relied most on a comparatively larger group of infrequent telic lexemes and also upon few – albeit very frequent – actionally underspecified lexemes.

5.2 The influence of L1 input

RQ₂ asked if frequency and contingency of perfective lexemes in L1 Italian might have affected the values of ΔP attraction and reliance of early L2 perfectives.

Table 5 reports the ten most frequent L2 perfectives occurring at early interviews. The first four positions of the table are occupied by lexically underspecified predicates. In addition, among the nine general-purpose predicates quoted by Theakston and colleagues (2004) and by Wulff et al. (2009), four (the equivalent of *done*, *gone*, *seen*, *come*) appear in the top ten list:

Table 5: The ten most frequent perfectives at early interviews.

Rank	Perfective	Token Frequency	English translation	Lexical aspect category
1	<i>fatto</i>	143	Done	Underspec
2	<i>andato</i>	111	Gone	Underspec
3	<i>visto</i>	108	Seen	Underspec
4	<i>detto</i>	98	Said	Underspec
5	<i>capito</i>	68	Understood	Telic
6	<i>arrivato</i>	59	Arrived	Telic
7	<i>venuto</i>	34	Come	Telic
8	<i>trovato</i>	29	Found	Telic
9	<i>parlato</i>	28	Talked	Atelic
10	<i>finito</i>	28	Finished	Telic

In order to compare L1 and L2 distribution of the sampled perfectives, we retrieved the normalized (per million) frequencies and the lexeme-morpheme associations of the sampled perfective predicates in four Italian corpora (online Appendix 4). By virtue of their different sizes, natures, and compositions, we assume the normalized occurrences of these four corpora to be a fair approximation of the written and spoken language input to which the L2 learners could have been exposed. To test whether L2 frequency matched the L1 input frequencies, we used the functions `chisq.pval` and `fisher.pval` from the R package “corpora” (Evert 2018). This function returns multiple *p*-values of Pearson’s chi-squared test and Fisher’s exact test for frequency comparisons from integer vectors. All 39 *p*-values from pairwise comparisons approached 0, suggesting that the frequencies from L1 and L2 corpora were not comparable. online Appendix 4 reports the normalized frequencies (per 1 million) of the 10 most frequent perfective predicates in the Corpus Pavia and in the four L1 corpora.⁹ The results show that most frequent L2 perfectives were underrepresented or inconsistently represented in the L1 rankings. To compare rankings beyond the

⁹ Occurrences were sorted out by CQL in SketchEngine, tag [tag = “VER:ppast”]. They comprise all inflected forms (masculine, feminine, singular, plural).

top 10 positions, we assigned each one of the top-10 L1 perfectives its actual ranking in the complete frequency list of the Corpus Pavia (Table 6).

Table 6: Rank comparison of the top 10 perfectives in L2 and L1 Italian corpora.

Verbs	English Translation	Pavia	ItTenTen	Clip	Lip	Perugia
<i>detto</i>	Said	1	2	2	1	2
<i>fatto</i>	Done	2	10	5	2	10
<i>andato</i>	Gone	3	1	6	5	1
<i>visto</i>	Seen	4	22	1	10	4
<i>capito</i>	Understood	5	17	35	22	22
<i>arrivato</i>	Arrived	6	14	58	17	3
<i>preso</i>	Taken	7	35	21	35	6
<i>trovato</i>	Found	8	60	59	14	35
<i>parlato</i>	Talked	9	7	72	12	5
<i>venuto</i>	Come	10	58	17	7	7

A Kendall’s τ correlation test showed no agreement between L1 and L2 rankings, except for the spoken corpus CLIP ($\tau = 0.55$, $p = 0.02$).¹⁰ This suggests that in general very frequent perfectives in L1 corpora were not so in the Corpus Pavia and vice versa, with the exception of a bunch of lexically underspecified, general-purpose verbs. In fact, among the 10 most frequent verbs, the most represented in all corpora were *detto* “said”, *fatto* “done”, and *preso* “taken”. Contextual relevance can explain the different distribution of perfectives in the corpus Pavia and in L1 corpora. For example, out of 386 instances of the lemma *capire* “understand”, 215 in the corpus Pavia are represented by two perfective forms: *ho capito* “I understood” and *hai capito?* “did you understand?” Such backchannel forms occur very frequently in clarification requests and are typical of native–nonnative interaction. To give another example, almost half of 334 total lexemes of the lemma *vedere* “see/look” in the corpus Pavia consist of two perfective forms: *ho visto* “I saw” and *hai visto* “you saw”. This verb was predicted to occur especially during retelling tasks based on video clips. Also, the abnormally high frequency of specific perfectives such as *sposato* “married” in the Corpus Pavia (ranked 22, frequency per million = 71) corresponded to a scene of *Modern Times* that learners were asked to retell. Such

10 As an anonymous reviewer correctly pointed out, the fact that the distribution of perfectives in the corpus Pavia correlates exclusively with the distribution of perfectives in a spoken Italian corpus is interesting and worth further investigation.

frequency does not reflect the frequency in L1 input (ranked 851, frequency per million = 0.42 in ItTenTen).

As to ΔP values, Figure 3 illustrates morpheme–lexeme attraction and reliance in the (normalized) L1 corpora.

Figure 3: Reliance and attraction in normalized L1 Italian data.

As expected – given the incomparably larger number of types, tokens, lexical entries and constructions in L1 corpora – attraction and reliance values of the same predicates were considerably lower in L1 corpora in comparison with the corpus Pavia. Beside the differences due to sizes, two other L1-L2 differences emerged: (a) the LA of predicates affected the values of reliance in L2 (Kruskal-Wallis chi-squared = 15.539, $df = 2$, $p = <0.001$, average between groups difference = 0.287) but not in L1 (Kruskal-Wallis chi-squared = 3.2996, $df = 2$, $p = 0.19$, average between groups difference = 0.03). In fact, while among the top ten reliable perfectives eight are telic, among L1 perfectives all actional classes are better balanced; (b) the values of attraction and reliance correlated in L1 corpora but not in early L2 data, meaning that – only in the L1 – the most/least reliable lexemes tended to be also those that attracted most/least the perfective. A similar correlation between attraction and reliance – that was absent at early interviews – appeared also in L2 data starting from mid interviews (see Section 6.3). To sum up, our data suggested that the values of ΔP attraction and reliance in early interviews did not reflect the contingency values of the same perfectives in L1 Italian.

5.3 ΔP changes across interviews

RQ₃ asked if lexeme–morpheme contingency scores changed over time. As table 7 and table 8 show, the ΔP scores of nearly all predicates in the sample declined from early to mid interviews, with the exception of *morto* “died”, *visto* “seen” and *finito* “finished”. On the contrary, in most cases the values of ΔP did not decline from mid to late interviews.

Table 7: Fluctuations of reliance values across interviews in the top-ten perfectives.

Verb	English	early	mid	late
<i>dimenticato</i>	Forgotten	0,828	0,545	0,649
<i>sbagliato</i>	Mistaken	0,743	0,390	0,397
<i>perso</i>	Lost	0,664	0,432	0,439
<i>finito</i>	Finished	0,645	0,345	0,284
<i>portato</i>	Brought	0,539	0,102	0,283
<i>morto</i>	Died	0,533	0,589	0,218
<i>dato</i>	Given	0,523	−0,005	0,016
<i>incontrato</i>	Met	0,474	0,263	0,38
<i>arrivato</i>	Arrived	0,469	0,318	0,267
<i>capito</i>	Understood	0,458	0,248	0,15

Table 8: Fluctuations of attraction across interviews in the top-ten perfectives.

Verb	English	early	mid	late
<i>fatto</i>	Done	0,101	0,002	−0,021
<i>visto</i>	Seen	0,093	0,090	0,026
<i>andato</i>	Gone	0,063	−0,011	−0,027
<i>detto</i>	Said	0,062	−0,024	0,001
<i>capito</i>	Understood	0,057	0,036	0,02
<i>arrivato</i>	Arrived	0,05	0,031	0,028
<i>venuto</i>	Come	0,025	−0,005	−0,012
<i>finito</i>	Finished	0,025	0,020	0,012
<i>trovato</i>	Found	0,022	−0,004	0,002
<i>imparato</i>	Learned	0,021	−0,003	−0,05

It is important to establish which altered value(s) in the contingency table (Section 4.4, Table 1) might have been most responsible for the decline in ΔP scores from early to mid interviews. As Table 9 shows, the relative frequency (density) of the perfectives compared to other verbal forms did not change from early to late interviews, nor did the width of the verbal lexicon (meant as number of different, non-perfective verbal lemmas, cell *d* of the contingency table).¹¹ Rather, the number of different perfective predicates increased significantly from early to mid interview period.

Table 9: Density of perfective forms in the verbal lexicon per interview.

Interview period	Total perfectives	Total verbs	Perfectives as fraction of total (%)	Number of different perfectives
Early	1,508	6,500	23	116
Mid	1,863	6,970	26	149
Late	2,169	8,639	25	145

This change may have been due to type frequency, which indicates an expansion of the repertoire of perfective forms available to the learner (the value in cell (c) of the contingency table).

As noted at the end of the previous section, the decreasing values of L2 attraction and L2 reliance became increasingly correlated over time in our data. Table 10 shows the changing relationship between attraction and reliance.

Table 10: Pearson’s correlation coefficients between attraction and reliance across interviews.

Interview	Pearson’s <i>r</i>	<i>p</i>
Early	0.10	0.34
Mid	0.60	<0.001
Late	0.70	<0.001

Starting from mid interviews, the more a lexeme attracted the perfective, the more the perfective relied on that lexeme, regardless of predicate frequency. Unlike early interviews, in late interviews telic lexemes displayed the highest values of both attraction and reliance. In Figure 4, ΔP values at late interviews are plotted.

¹¹ Vocabulary size increased dramatically from mid to late interviews.

Figure 4: Reliance (y) and attraction (x) at late interviews.

To sum up, our data showed that: (a) lexeme–morpheme contingency scores decreased over time, as learners’ proficiency increased; (b) ΔP changes over time were likely due to the increasing repertoire of perfective forms (type frequency) available to learners; (c) attraction and reliance became correlated over time; (d) in late interviews, telic lexemes displayed the highest values of both attraction and reliance.

6 Learners’ L1s, proficiency, and length of residence

Learners’ L1s, length of residence in Italy at first interview, proficiency (as it was established by the authors of the corpus) did not influence the ΔP scores, the overall number of verbs produced, the number of perfectives produced in each interview, or the percentage of perfectives among the verbal forms (all $p \geq 0.15$ at monofactorial ANOVA and Kruskal-Wallis tests, respectively, for parametric and non-parametric outcome variables). Proficiency did influence the overall number of perfectives produced ($\chi^2 = 9.6$, $p = 0.02$). However, basic and advanced learners patterned alike, possibly confirming either that the criteria used for establishing proficiency were not reliable or that the production of perfectives alone strongly depended on the topic of interviews or on the elicitation task. Finally, the number of perfectives produced in each interview and the percentage of perfectives out of the overall number of verbs correlated negatively with

the type/token ratio of verbal forms ($r = -0.55$, $p = 0.006$). This increase was both expected – given the expansion of type frequency over time – and generalized across learners, L1s, length of stay, and proficiency level being all irrelevant factors (all $p \geq 0.1$).

7 Summary, theoretical implications and open issues

As to RQ₁, the emergence and productivity of the perfective morpheme in the corpus Pavia was contingent upon a large number of infrequent telic lexemes (such as *sbagliato* “mistaken”, *dimenticato* “forgotten”, *perso* “lost”, *finito* “finished”, *arrivato* “arrived”), but also upon a bunch of very frequent, actionally underspecified, general purpose ones (e.g. *preso* “taken”, *dato* “given”, *visto* “seen”). What such lexemes have in common beyond LA was that most of them were all highly expected given the elicitation tasks and the topics of interviews utilized in the corpus Pavia. Unlike reliance, the values of lexeme-morpheme attraction were negligible across the board. As to RQ₂, L1 frequency and distribution seemed to play no role in the emergence of the L2 perfectives in our data.¹² As to RQ₃, ΔP values decreased consistently over time as learners’ proficiency increased. Finally, as interviews progressed, perfective predicates reached a target-like configuration where attraction and reliance became correlated.

The results described above may have implications for the theoretical issues raised in Sections 1 and 2 of the current paper, but their interpretation is difficult and they seem to raise more questions than they provide answers. Certainly, on the one hand the emergence of the perfective morpheme in the corpus Pavia was contingent on telic predicates. *Related* CL (Section 2.4) and the LAH in fact predicted that the expansion of perfective types begins with semantically coherent (telic) lexemes. On the other hand, highly reliable telic predicates were not frequent and their attraction values remained negligible. The dissociation

¹² As an anonymous reviewer pointed out, the general lack of the similarity between the distributional pattern of perfective morphemes in L1 and those in L2 can be due to the characteristics of tasks and the nature of NS-NNS communication. In that case, it should not be claimed that the frequency and distribution pattern in L1 plays no role in L2. I totally agree, to the extent that this observation casts doubts not only to my data, but also on many studies based on corpora comparison (Section 5.2) and on the generally accepted claim that we can compare learner corpora and native corpora to test how learners’ exposure to the L1 input affect L2 acquisition.

between attraction and reliance values of early telic perfectives needs to be explained, together with the fact that the most attracting lexemes were not telic, but actionally underspecified (and very frequent). As to the role of L1 frequency, our results differed greatly from those reported by Tracy-Ventura and Cuesta Medina (2018) and Wulff et al. (2009) (Section 2.6). In fact, the early L2 perfectives in our data were neither the most frequent nor the most strongly attracted to the perfective construction. Our data suggested that attraction – but not reliance – is *distinctive*, meaning that the distribution of its values was Zipfian and characterized a limited set of actionally underspecified lexemes (see Figure 2). Finally, the results also indicated that *transiency* is a characteristic of CL. Decline in association scores and their gradual correlation marked the end of CL, which in turn corresponded with the spread of the perfective morpheme in learners' developing grammar. The decline of CL was generalized across both predicates (regardless of LA) and learners (regardless of L1). Our data confirmed that type frequency and ΔP (cue–outcome association) correlated inversely. Decreasing values of association-contingency scores in our study might mean that learners were becoming aware that the construction featured a [+perfective] trait having its own particular distribution, which did not correspond to that of the lexeme. The statistical analysis adopted in this study – and in particular ΔP – is replicable to other domains where contingency of stem-affix alternations may provide cues for observing the developing L2 grammar.

There are at least three issues that cannot be addressed in this study. First, although the ratio between perfective forms and verbal lemmas was very similar across periods of interviews, the comparison of ΔP scores from different datasets is problematic. We have seen that ΔP scores declined over time, but we ignore whether and to what extent such fluctuations can be deemed “significant” or not. We also ignore how this significance could be evaluated. When comparing proportions of occurrences from different datasets, it is difficult to draw any conclusions without utilizing measures of significance that can keep aside sample size and effect size (see Gries 2019). The second issue concerns whether aggregated longitudinal data can be still considered as longitudinal data rather than cross-sectional. According to some, truly longitudinal data set is one where there is data points from one or more learners at multiple time points, allowing the researcher to track an individual learner's development over time. One can reply that the temporal dimension of language development characterizes also “aggregated longitudinal data”. Future research should integrate the analysis of fluctuations of aggregated ΔP scores with the analysis of fluctuations in individual data. Finally, it should be explained why we could not find a significant relationship between the length of residence in Italy at the first interview and ΔP whereas we did find a correlation between ΔP and the learners' longitudinal

development. If this latter factor led to lower ΔP values, we could have observed the same between the learners' different length of residence in Italy. It is of course possible that language development is not simply a matter of time passage and that hidden or nested variables – other than time – affected the results.

References

- Andersen, Roger. 2002. The dimensions of “Pastness”. In Rafael Salaberry & Yasuhiro Shirai (eds.), *The L2 acquisition of tense-aspect morphology*, 79–106. Amsterdam-Philadelphia: John Benjamins.
- Andersen, Roger & Yasuhiro Shirai. 1994. Discourse motivations for some cognitive operating principles. *Studies in Second Language Acquisition* 16(2). 133–156.
- Andersen, Roger & Yasuhiro Shirai. 1996. The primacy of aspect in first and second language acquisition: The pidgin/creole connection. In C. Ritchie & T. K. Bhatia (eds.), *Handbook of second language acquisition*, 527–570. New York, NY: Academic Press.
- Beckers, Tom, Jan De Houwer & Helena Matute. 2007. Editorial: Human contingency learning. *The Quarterly Journal of Experimental Psychology* 60(3). 289–290.
- Bertinetto, Pier Marco. 2001. On a frequent misunderstanding in the temporal-aspectual domain: The ‘Perfective = Telic Confusion’. In C. Cecchetto, G. Chierchia & M. T. Guasti (eds.), *Semantic Interfaces [Reference, Anaphora and Aspect]*, 177–210. Stanford: CSLI Publications.
- Bley-Vroman, Robert. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(1). 1–17.
- Champollion, Lucas & Manfred Krifka. 2016. Mereology. In P. Dekker & M. Aloni eds., *Cambridge handbook of formal semantics*. Cambridge: Cambridge University Press, 513–541.
- Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- Croft, William & Allan Cruse. 2004. *Cognitive linguistics*. Cambridge–New York: Cambridge University Press.
- Dowty, David. 1979. *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Ellis, Nick. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24.
- Ellis, Nick. 2012. Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics* 32. 17–44.
- Ellis, Nick. 2016. Cognition, corpora, and computing: triangulating research in usage-based language learning. *Language Learning* 67(51). 40–65.
- Ellis, Nick & Fernando Ferreira-Junior. 2009. Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal* 93(3). 370–385.
- Evert, Stefan. 2018. Package ‘corpora’. <http://SIGIL.R-Forge.R-Project.org/>.
- Giacalone Ramat, Anna (ed.). 2003. *Verso l'italiano. Percorsi e strategie di acquisizione*. Roma: Carocci.

- Gluck, Mark & Gordon Bower. 1988. From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General* 117(3). 227–247.
- Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Gries, Stefan. 2015. Statistics for learner corpus research. In G. Gilquin, S. Granger & F. Meunier (eds.), *The Cambridge handbook of learner corpus research*, 159–182. Cambridge: Cambridge University Press.
- Gries, Stefan. 2018a. Mechanistic formal approaches to language acquisition. Yes, but at the right level(s) of resolution. *Linguistic Approaches to Bilingualism* 8(6). 733–737.
- Gries, Stefan & Nick Ellis. 2015. Statistical measures for usage-based linguistics. *Language Learning* 65(Supplement 1). 1–28.
- Gries, Stefan & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on ‘alternations’. *International Journal of Corpus Linguistics* 9(1). 97–129.
- Gries, Stefan & Sabine Stoll. 2009. Finding developmental groups in acquisition data: Variability-based neighbour clustering. *Journal of Quantitative Linguistics* 16(3). 217–242.
- Gries, S.T. 2019. 15 years of collostructions: some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). In S. Hunston and F. Perek (eds.) *Constructions in Applied Linguistics, Special Issue of International Journal of Corpus Linguistics* 24(3), 385–412.
- Hommel, Bernhard & Wolfgang Prinz (eds.). 1996. *Theoretical issues in stimulus-response compatibility*. Amsterdam: North-Holland.
- Hommel, Bernard, J. Müsseler, G. Aschersleben & Wolfgang Prinz. 2001. The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences* 24. 849–878.
- Klein, Wolfgang. 1994. *Time in language*. London: Routledge.
- Krifka, Manfred 1992. A compositional semantics for multiple focus constructions. Proceedings of Semantics and Linguistic Theory (SALT) 1. *Cornell Working Papers in Linguistics*, 10. 127–158.
- Lardiere, Donna. 2003. Revisiting the comparative fallacy: A reply to Lakshmanan and Selinker, 2001. *Second Language Research* 19(2). 129–143.
- Li, Ping & Yasuhiro Shirai. 2000. *The acquisition of lexical and grammatical aspect*. Berlin-New York: Mouton De Gruyter.
- Mourelatos, Alexander. 1978. Events, processes, and states. *Linguistics and Philosophy* 2(3). 415–434.
- Nosofsky, R. M. 2014. The generalized context model: An exemplar model of classification. In M. Pothos & A. Wills (eds.), *Formal approaches in categorization*, 18–39. New York: Cambridge University Press.
- Prinz, Wolfgang. 2018. Contingency and similarity in response selection. *Consciousness and Cognition* 64. 146–153.
- Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny & Kirsten Thorpe. 2010. The effect of feature label order and their implications for symbolic learning. *Cognitive Science* 34(6). 909–957.
- Rastelli, S. 2008. A compositional account of L2 verb actionality and the aspect hypothesis. *Lingue e Linguaggio* 7. 261–289.
- Reeder, P. A., E. L. Newport & R. N. Aslin. 2010. Novel words in novel contexts: The role of distributional information in form-class category learning. In S. Ohlsson & R. Catrambone

- (eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, 2063–2068. Austin, TX: Cognitive Science Society.
- Salaberry, Rafael & Llorenç Comajoan (eds). 2013. *Research design and methodology in studies on L2 tense and aspect*. Boston, MA & Berlin: De Gruyter Mouton.
- Shanks, David. 2007. Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology* 60(3). 291–309.
- Smith, Carola. 1997. *The parameter of aspect*. Dordrecht: Springer.
- Stefanowitsch, Anatol Forthcoming. *Corpus linguistics: A guide to the methodology*. Freie Universität Berlin: Language Science Press. <http://langsci-press.org/catalog/book/148>.
- Stoll, Sabine & Stefan Gries. 2009. How to measure development in corpora? An association strength approach. *Journal of Child Language* 36. 1075–1090.
- Theakston, A. L., E. V. M. Lieven, J. M. Pine & C. F. Rowland. 2004. Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language* 31(1). 61–99.
- Tracy-Ventura, N. & J. A. Cuesta Medina. 2018. Can native-speaker corpora help explain L2 acquisition of tense and aspect? A study of the “input”. *International Journal of Learner Corpus Research* 4(2). 277–300.
- Tversky, A. & I. Gati. 1978. Studies of similarity. In E. Rosch & B. Lloyd (eds.), *Cognition and categorization*, 79–98. Oxford, England: Lawrence Elbaum Associates.
- Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca: Cornell University Press.
- Verkuyl, Henk. 1993. *A theory of aspectuality*. Cambridge: Cambridge University Press.
- Viberg, Åke. 2002. Basic verbs in lexical progression and regression. In P. Burmeister, T. Piske & A. Rohde (eds.), *An integrated view of language development: Papers in honor of Henning Wode*, 109–134. Trier, Germany: Wissenschaftlicher Verlag Trier.
- Wulff, S., N. Ellis, K. Bardovi-Harlig, C. J. Leblanc & U. Römer. 2009. The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal* 93(3). 354–369.
- Yang, Charles. 2018. A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism* 8(6). 665–706.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cllt-2019-0071>).

Bionote

Stefano Rastelli

Stefano Rastelli is senior lecturer in SLA at the University of Pavia (Italy). He published books and articles on Statistical Learning, L2 acquisition of syntax and on the neural correlates of the teaching/acquisition interface.