L2 language development in oral and written modalities

Myeongeun Son

Department of English Language and Literature, Gangneung-Wonju National University, Gangneung, Gangwon Province, South Korea
Email: sonmyeo1@gwnu.ac.kr.

Abstract

This study investigates whether second language (L2) learners' language development and accuracy in production are comparable across oral and written modalities on the basis of Pienemann's processability theory (PT). Eighty-seven English as a Foreign Language (EFL) learners, from high beginner to advanced levels, completed comparable speaking and writing tasks designed to elicit particular morphosyntactic structures predicted by PT to correspond to L2 stages of development. Time constraints encouraged participants to respond spontaneously, thus drawing on implicit knowledge. Implicational scaling shows correlations that suggest comparable language development between the modalities. However, accuracy was higher earlier in the written than in the oral modality, and accuracy in the written modality was more stable. The results provide a clearer understanding of the similarities and differences of L2 oral and written development and demonstrate that PT can be applied to L2 writing.

Introduction

Speaking and writing have different characteristics that affect how second language (L2) learners engage with them. One important difference is that speaking requires more spontaneous production, while writing provides more opportunities to revise and edit, because it typically entails limited simultaneous interaction (Byrnes & Manchón, 2014; Gilabert, Manchón, & Vasylets, 2016; Williams, 2012). The distinct features of each modality often lead L2 researchers to consider them separately.

The acquisition of the two modalities, however, is not completely separate. Nevertheless, most L2 development studies analyze only oral production, and L2 development in writing is less well theorized, with most writing research instead focusing on topics such as learners' changes in knowledge of genres, goal setting, and the writing process (Cumming, 2012; Sasaki, 2004; Tardy, 2012; also see Polio, 2017; Polio & Park, 2016). To begin to address this gap and given the fundamental connection between speaking and writing, this study explores whether processability theory (PT) (Pienemann, 1998, 2005), which offers a framework for understanding L2 development based on oral production, can be applied to L2 written production. PT

posits a universal implicational order of L2 development. According to PT, as learners develop the L2, they acquire processing procedures at a higher level, which allows more diverse structures to emerge.

This study compares the emergence of PT's target morphosyntactic structures in the oral and written production of learners at four proficiency levels. It further compares the learners' accuracy in their oral and written production of the structures. Comparing the emergence and accuracy of particular morphosyntactic structures between oral and written production will lead to better understanding of the similarities and differences of language development in the two modalities.

Literature review

Modality

In terms of constraints on online processing, speaking and writing typically differ along three dimensions: (a) the presence or absence of an audience during production, (b) the stability of the language signal, and (c) the degree of control of the language user over the linguistic output (Ravid & Tolchinsky, 2002, p. 426). Speaking is usually, but not always, interactive and writing is usually, but not always, not interactive. In particular, speakers usually produce speech while interacting with listeners, whereas writers do not usually expect immediate interaction with readers while producing writing. The lack of immediate interaction also means that writing is usually completed at the writer's own pace, which is much slower but more stable than the pace of spontaneous speech production (Gilabert et al., 2016). In addition, written production shows less variation in accuracy than oral production, because writers can usually take time to revise and edit their writing before the audience reads it (Granfeldt, 2008). Writers thus generally have more control over their attentional resources and more opportunities to pay attention to their production, compared to speakers (Byrnes & Manchón, 2014; Williams, 2012). Due to these distinct characteristics, writing, even under time constraints, allows access to both explicit (declarative) and implicit (procedural) knowledge, while speaking frequently demands the use of implicit (procedural) knowledge for fluent execution (Polio, 2012).

Speaking and writing have these different characteristics, regardless of the status of the language (e.g., native language [L1], L2) for the language user. In this regard, some differences between L2 speaking and L2 writing have been found in the field of L2 acquisition. For instance, similar to L1 speakers, L2 learners can attend to form more in writing than in speaking, which in turn may lead to specific structures and vocabulary appearing earlier in L2 learners' writing than in their speaking (Polio, 2012). Despite their different characteristics, research has shown overlap in L2 oral and written production in terms of cognitive processes; complexity, accuracy, and fluency (CAF); and developmental sequence.

Cognitive processes in speaking and writing

Theoretical approaches to the cognitive processes involved in L2 speaking and writing either try to explain one or the other, such as Levelt's (1989) speaking model (conceptualization, formulation, and articulation) and Kellogg's (1996) writing model (formulation, execution, and monitoring), or suggest that speaking and writing use the same cognitive processes but draw on different subsets of these processes (Bourdin & Fayol, 1994, 2000; Brown, McDonald, Brown, & Carr, 1988). In particular, although both Levelt's speaking model and Kellogg's writing model include a formulation stage

in which language users generate acceptable forms and produce a temporary linguistic output, Kellogg's monitoring stage in the writing process allows language users to go through and edit their writing. Compared to the stages of the speaking process, the stages of the writing process are more interactive and recursive (Kellogg, 2001; McCutchen, 1996). The later stages of the writing process may be positively or negatively affected by the earlier stages, and writers can go back and forth among these stages until they complete their production. In addition, Brown et al. (1988) argued that language production includes formulation, execution, and monitoring stages, regardless of modality. They posited that formulation processes might be interchangeable across modalities but would be combined with different sets of execution processes and different monitoring processes to produce either oral or written language.

Regardless of which approach it takes, existing research often tacitly assumes that speaking and writing production share at least some of the cognitive processes. If this is the case, then some characteristics of speaking production may also be found in writing production, and vice versa.

CAF in oral and written production

Task-based language teaching studies that touch on the effects of task modality on performance have had mixed results (Gilabert et al., 2016; Granfeldt, 2008; Kormos & Trebits, 2009; Kuiken & Vedder, 2011; Son, 2022; Tavakoli, 2014; Vasylets, Gilabert, & Manchón, 2017; Zalbidea, 2017). For instance, Kuiken and Vedder (2011) and Zalbidea (2017) both compared an L2 oral task group to an L2 written task group. Kuiken and Vedder found greater syntactic complexity in the written task (i.e., writing a letter) than the oral task (i.e., leaving a voice message). In contrast, Zalbidea's study found greater complexity in the oral task (i.e., leaving a voice message) than the written task (i.e., writing an email). Other studies have observed no significant difference in syntactic complexity between oral and written tasks (Granfeldt, 2008; Kormos, 2014; Kormos & Trebits, 2009; Vasylets et al., 2017). Likewise, some studies (Kormos, 2014; Kormos & Trebits, 2009; Tavakoli, 2014; Zalbidea, 2017) showed greater accuracy in a written task; they suggested that the additional opportunities to revise and monitor in a written task allow for greater accuracy. On the contrary, other studies had different results (Granfeldt, 2008; Vasylets et al., 2017).

Changes in CAF over time have also been compared between oral and written modalities (Bulté & Housen, 2009; Serrano, Tragant, & Llanes, 2012; Weissberg, 2000), providing evidence of how L2 knowledge develops (e.g., Larsen-Freeman, 1978). For example, Serrano et al.'s (2012) longitudinal study implemented oral narrative tasks and descriptive essay tasks three times during a 1-year study abroad program. The study showed that although both oral and written production developed, the time at which each construct in CAF developed differed across the two modalities. For instance, fluency and lexical diversity in oral production developed earlier, while accuracy in both modalities developed later.

L2 grammar development

The development of learners' grammatical knowledge seems comparable across speaking and writing to some extent (e.g., Boss, 2008; Byrnes & Sinicrope, 2009; Kyle, Crossley, & Verspoot, 2021; Park, 2017; Weissberg, 2006; Zalbidea, 2021). For instance, Zalbidea (2021) examined the development of beginner-level L2 Spanish learners' knowledge of future and clitic forms with a pre/post/delayed-posttest experimental design. The study found that after the treatment, speaking and writing groups had

comparable scores, showing significant gains, on the tasks, although the writing group showed better performance on some parts (i.e., in a production task on the posttest and on the clitic in an aural acceptability judgment task on the delayed posttest).

In addition, many studies posit an L2 acquisition order of specific morphosyntactic forms (e.g., Gass, 1980). Most of these investigations discuss only oral production, while only a few address written production (e.g., Boss, 2008; Byrnes & Sinicrope, 2009; Kyle et al., 2021; Park, 2017; Weissberg, 2006). However, the results of some recent L2 writing studies indicate the possibility of applying L2 speaking studies' approach to writing.

Boss (2008) investigated the acquisition order of L2 German verb morphology, which Pienemann (1998) originally predicted based on oral production, in the written modality. The results indicated that as L2 learners' knowledge of German developed, the number of present tense conjugations of *sein* and *haben* that emerged increased, aligning with Pienemann's predictions, while the emergence of past participles of irregular verbs did not appear to fit Pienemann's predictions. Likewise, Byrnes and Sinicrope (2009) explored whether a universal progression for relative clauses suggested by the Noun Phrase Accessibility Hierarchy (Keenan & Comrie, 1977) applies to L2 written production. They found that most of the learners' use of relative clauses followed the Noun Phrase Accessibility Hierarchy; however, 20–30% of the relative clauses in the data contradicted the predictions.

Both Boss (2008) and Byrnes and Sinicrope (2009) suggested that the learning process in the written modality contrasts with the learning process in the oral modality to some extent. However, caution is needed in comparing the results from these studies to the results of studies on learners' development in speaking. As Norris and Ortega (2003) pointed out, researchers attempting to assess the emergence of complex phenomena must be careful not to base their interpretations "on a lack of evidence, as opposed to evidence for the lack of emergence…[I]t is likely that measurement data are more frequently underinterpreted when researchers do not adequately conceptualize the complexities of measurement behaviors that they intend to elicit" (p. 733–734). That is, a lack of evidence for similarities across the modalities may be due not to different developmental paths in oral and written modalities but to research design.

For example, Boss (2008) chose specific topics for the writing tasks (e.g., a travel diary) to elicit particular structures; but the number of tasks (one or two) and the length of writing (around 100 words) for each task may not have been enough to elicit sufficient past participles of weak verbs in German (e.g., *sagen → gesagt*, [say → said]). In addition, the instructions for the writing tasks allowed the learners to avoid using certain structures. Likewise, Byrnes and Sinicrope (2009) analyzed tasks designed to evaluate learners' overall achievement at the end of a course. Writing tasks designed specifically for the elicitation of target structures might better reveal a developmental progression for written production that could be validly compared to the developmental progression of oral production.

In sum, very few studies have investigated the L2 developmental sequence in the written modality, and those few have not been based on the theoretical frameworks widely used to explore L2 development in the spoken modality. Nevertheless, some similar patterns have been observed. But because external factors (e.g., task type, instructions) may have led to findings that show differences between development in oral production and development in written production, it is important to investigate whether particular developmental theories can be applied to both modalities with the use of similar tasks (Polio, 2012, 2017).

Although there is no consensus on the effects of modality, previous studies have found comparable effects across modalities at least to some extent. In addition, only a

few studies have attempted to apply theoretical frameworks of L2 development to the written modality, and they have had some methodological limitations (e.g., Boss, 2008; Byrnes & Sinicrope, 2009). Given the connections between the oral and the written modalities and the possibility of the application of theoretical frameworks to both, this study casts light on whether L2 language development is consistent across the oral and the written modalities on the basis of PT, which allows for investigating theory-based L2 development in writing as well as comparability of L2 development between the two modalities (Polio, 2012 2017; Polio & Park, 2016).

Processability Theory

Pienemann (1998) proposed PT to explain the developmental stages in language learners' interlanguage. PT hypothesizes that, at each developmental stage, language learners acquire a processing procedure that manages information transfer, which is necessary for the production of morphosyntactic structures. The processing procedures are based on lexical functional grammar's (Kaplan & Bresnan, 1982) feature unification process, which posits three processing procedures, in which language users (a) classify the grammatical information in a lexical item, (b) store the information temporarily, and (c) utilize the information at different points in the constituent structure. Pienemann (1998, 2005) argued that processing procedures are hierarchical and the levels of the hierarchy can be addressed through feature unification. In addition, Pienemann claimed that language learners go through these developmental stages in a universal implicational order. In other words, a learner cannot skip lower stages and follow a different developmental order. This view of development in processing procedures is based on Levelt's (1989) speaking model, which suggests that language processing is incremental and linear: While surface forms are still being constructed, grammatical information is temporarily stored in grammatical memory storage, which ultimately allows automatic processing of the information. According to PT, as language learners reach higher developmental stages, the feature unification (or matching) described by lexical functional grammar takes place across greater structural distances within sentences.

Although PT research has primarily analyzed L2 learners' oral production, the hierarchy of processing procedures is not limited to a specific modality or specific task type. Pienemann's (1998) steadiness hypothesis posits that the developmental order does not change across different communicative tasks as long as the tasks are designed to draw on the same production skill (e.g., Ellis, 2008; Pienemann, 1998; Spinner, 2011). For instance, Pienemann (1998) found that six different communicative tasks reflected the developmental order of PT in learners' production, supporting the steadiness hypothesis, although morphology production showed a few mismatches (i.e., in about 1% of the data).

Based on the processing procedures that become available to learners at each stage, PT proposes six developmental stages: Lemma access (Stage 1), Category procedure (Stage 2), Phrasal procedure (Stage 3), Verb phrase procedure (Stage 4), S-procedure (Stage 5), and Subordinate clause procedure (Stage 6). In Stage 1, learners are limited to producing single words or formulaic expressions. Lacking processing procedures, they are not ready to exchange information, so feature matching or unification does not occur. At this stage, nonlinguistic strategies, such as facial expressions, play a crucial role in keeping communication going. In Stage 2, when category procedures become available, learners are able to assign categorical information to lexical items. At this

stage, operations occur only within a single constituent (e.g., *dog[s]*). When it comes to syntax, learners start to produce syntactic structures with canonical word order (e.g., *Mary woke up at 7:00 am*).

Next, in Stage 3, a single lexical item that involves categorical information can take the role of head of a phrasal category. Within a phrasal category, feature information can be exchanged across elements (e.g., *these dogs*). In terms of syntax, some fronting elements, such as adverbs and prepositional phrases, emerge (e.g., *Yesterday, Mary woke up at 7:00 am*). In addition, learners begin to produce questions starting with *do*-verbs (e.g., *Did Mary wake up at 7:00 am?*). According to the topic hypothesis (Pienemann, Di Biase, Kawaguchi, & Håkansson, 2005; Pienemann & Lenzing, 2020), at this stage learners also start to differentiate syntactic topics and grammatical functions from the subject.

In Stage 4, learners can share a feature across phrases (e.g., *Mary woke up at 7:00 am and came to the office at 9:00 am*). In terms of syntax, learners begin to produce yes-no questions by inverting the subject and auxiliary verb, and to produce a *wh*-question with copular verbs or other auxiliary verbs (e.g., *Has Mary met you?; where is Mary?*). In Stage 5, grammatical features are matched across phrases. Subject and verb agreement of number and person features occurs at this stage (e.g., *Mary loves these dogs*). Learners also start to produce *wh*-questions in which *do* or other auxiliary verbs are in second position (e.g., *Where did you find this book?*). Finally, in Stage 6, the processing procedures for subordinate clauses are developed, such as *wh*-complement clauses in English (e.g., *Mary shows what she made yesterday*). A grammatical *wh*-complement in English does not include auxiliary inversion, which is necessary to make a question. When learners reach this developmental stage, they can differentiate the syntactic orders of a *wh*-complement and a *wh*-question and produce the *wh*-complement with the correct grammatical order.

In short, as processing procedures develop, the morphosyntactic structures that a learner can produce are more various. The processing procedures and morphosyntactic structures that become available to learners at each developmental stage are summarized in Table 1.

PT has been confirmed by many studies testing several languages (e.g., Bonilla, 2015; Jansen, 2008) in addition to English (e.g., Pienemann, 1998; Spinner, 2011). Although the studies had participants with diverse L1 backgrounds, different age ranges, and different contexts of language acquisition (i.e., foreign language vs. L2; classroom

Table 1. Processing Procedures Applied to English (Pienemann, 2005, p. 24)

Stage	Processing procedure	L2 process	Morphology	Syntax
6	Subordinate clause procedure	Main and subordinate clause		Cancel inversion (INV)
5	S–procedure	Interphrasal information	SV agreement	*Do*2nd, AUX2nd
4	VP–procedure	Interphrasal information	Tense agreement	Y/N inversion, copula inversion
3	Phrasal procedure	Phrasal information	NP agreement, neg+V	Adv, *do*–fronting, topicalization
2	Category procedure	Lexical morphology	Plural, possessive pronoun	Canonical order
1	Word/lemma	"words"	Invariant forms	Single constituent

vs. naturalistic setting), their results lend support to the implicational stages that PT proposes. In other words, the stages appear to be applicable to typologically different languages.

Accuracy and PT

The criterion of development in PT is the emergence of structures, so accuracy is not a primary focus. According to Pienemann (1998, p. 132), different morphosyntactic structures follow distinct developmental paths to reach accuracy; in addition, the rate at which the production of various structures reaches specific levels of accuracy varies. Therefore, Pienemann argued that accuracy, unlike emergence, is not a valid measurement of language learners' development.

However, the PT perspective on language learners' variability in production and automatization provides an opportunity to measure their accuracy in the use of morphosyntactic structures together with emergence, for the following reasons. When language learners attempt to produce morphosyntactic structures at stages higher than their present stage of development, they display inaccuracy and create variations (e.g., *What Mary do?* instead of *What does Mary do?*), because they have not yet developed the required processing procedure to produce the structures (Pienemann, 1998, 2005). In other words, based on a learner's developmental stage, his/her interlanguage has structurally limited variations: Learners initially produce inaccurate morphosyntactic forms and only later, when the relevant processing procedures are developed, become able to produce accurate forms.

PT defines language acquisition as acquisition of procedural knowledge that learners use to process their L2. In addition, developing processing procedures leads to automatization (Pienemann, 1998, 2005; Pienemann & Håkansson, 1999). When the underlying processing becomes automatized, a specific structure can be produced with procedural knowledge rather than metalinguistic knowledge (Kawaguchi & Di Biase, 2012). Acquisition of a certain structure includes acquisition of the routine that makes the processing procedure of the structure available. Research has shown that such automatization of the underlying routine is closely related to increased accuracy in producing the structure as well as processing efficiency. For instance, Anderson (1992) demonstrated that L2 learners show faster production and fewer errors when they reach the procedural and automatic stages for a given structure. In addition, according to Hulstijn, Van Gelderen, & Schoonen (2009), coming across novel vocabulary leads L2 learners to establish new form-meaning connections within their mental lexicons, while coming across familiar words reinforces pre-existing form-meaning connections. Through these processes, the learners experience automatization and become able to perform more quickly and accurately.

Nevertheless, until a routine is automatized for particular learners, they may show variability in their production of the morphosyntactic form. Even at the same developmental stage, their accuracy with different structures may differ based on the extent to which the underlying processing is automatized. In short, while accuracy is not a measure that PT utilizes, there is leeway within the PT framework to consider accuracy together with emergence.

Lee and Spinner (2017) conducted elicited oral production interviews with L2 learners. They measured accuracy as well as the emergence of target structures at different stages. Accuracy was calculated by means of obligatory occasion analysis (e.g., Ellis & Barkhuizen, 2005), and accuracy of all the target structures that each learner produced was measured, regardless of occurrence of emergence. The study showed that

accuracy differed across participants and morphological structures. For instance, participants who had reached Stage 5 had different accuracy rates (from 30% to 91%). Participants at Stage 3 had substantially lower accuracy for Third person singular -s than those at Stage 5. In addition, within Stage 2, Possessive 's had the highest accuracy (77%), whereas Past -ed had the lowest accuracy (51%) among the four structures analyzed. These results may be due to the different extents to which the underlying routine of the forms had become automatized for different learners. Although the study targeted only several morphological forms, it cast light on the possibility of using the PT framework to investigate accuracy and emergence simultaneously to gain understanding of language learners' development.

PT in the written modality

Pienemann's (1998) steadiness hypothesis posits that L2 learners' ability to process morphosyntactic structures is consistent across different oral tasks, if these tasks require the same knowledge and skills. A few studies have examined the topic in written production. For example, in Håkansson and Norrby's (2007) study, 20 L1-English students of L2 Swedish completed free writing and English-to-Swedish translation tasks. Oral data were also collected through communicative tasks designed to elicit particular morphosyntactic structures.

The results indicated that participants' development in written production was parallel to their development in oral production. It is important to note that there was no time constraint in the writing tasks. Thus, unlike spontaneous speech production, the writing tasks allowed the participants to take additional time to use their declarative knowledge and to plan and monitor their production in writing. Despite the difference, participants did not produce higher level morphosyntactic structures in writing. However, the participants' oral and written production were different in terms of language complexity. For instance, although the participants met the criterion of emergence of subordinate clauses in both the speaking and the writing tasks, they tended to avoid subordinate clauses in the speaking tasks. Based on the results, Håkansson and Norrby (2007) concluded that L2 learners' ability to process structures may be stable, as predicted by PT, but that the complexity they produce may differ across the modalities.

Håkansson and Norrby's (2007) study was the first to attempt to investigate whether PT can be applied to L2 learners' written production. However, some elements of the study suggest that the results should be interpreted cautiously and that further research is needed. First, because the participants all were at fairly high levels of development, it is difficult to conclude that processability is identical across the two modalities based on their production. In addition, one of the written tasks was a translation task. This task may not be as burdensome as other communicative tasks, as translating does not require learners to develop ideas and generate morphosyntactic structures based on their ideas. Rather, the learners needed only to translate "pre-made" English sentences into Swedish. Given differences in task difficulty, written production in a translation task may not reflect processability well. Employing similar tasks for speaking and writing would provide a better understanding of the application of PT to the written modality.

The possibility of applying PT to writing remains an open question due to the scarcity of research attempting to do so. In addition, some of the different results across the modalities may be due to methodological weaknesses, such as the use of noncomparable task types. To fill this gap, this study hypothesizes that the L2 developmental

sequence is comparable across the oral and written modalities based on their several overlapping characteristics and assesses whether PT predicts the order of emergence for written production as well as oral production by employing comparable oral and written tasks (see also, Polio, 2012, 2017). In addition, given PT's perspective on learners' variability and automatization in their production and inherent differences in the oral and written modalities (i.e., additional opportunity to revise and edit in writing), this study also explores developmental sequences across the modalities.

Research Questions:

1. Does the emergence of morphosyntactic structures in L2 learners' oral and written production follow the developmental order predicted by PT?
2. Does accuracy in L2 learners' oral and written production develop in parallel with the development of emergence of morphosyntactic structures?
3. Does task modality affect the development of accuracy?

Method

Participants

The study's participants were 87 L1-Korean EFL learners (47 female; age: $M = 22.97$, $SD = 2.53$), all undergraduate students who had learned English in instructional settings but rarely used the language elsewhere. Thirty had participated in various work/study-abroad programs in English-dominant countries (years: $M = .31$, $SD = .76$). The participants were categorized into four proficiency levels (A2–C1 in the Common European Framework of Reference for Languages [CEFR]) based on their standardized test scores (e.g., Test of English as a Foreign Language [TOEFL], International English Language Testing System [IELTS], Test of English for International Communication [TOEIC]; see Papageorgiou Tannenbaum, Bridgeman, & Cho, 2015 for CEFR equiv-alencies). The four proficiency groups were high beginner (A2, $N = 21$), low interme-diate (B1, $N = 22$), high intermediate (B2, $N = 22$), and advanced (C1, $N = 22$).

Materials

Speaking tasks

Each of the 11 speaking tasks was conducted one-on-one between the learner and me, and each was designed to elicit a particular morphosyntactic structure that charac-terizes a stage in PT. The target structures in the tasks were S neg V, Plural -s, Progressive -ing, Possessive 's, and Past -ed in Stage 2; Plural NP in Stage 3; and Third person singular -s in Stage 5; and Cancel INV in Stage 6. (Following previous studies, e.g., Di Biase, Kawaguchi, & Yamaguchi, 2015, multiple Stage 2 morphosyntactic structures were included to enable an investigation of a developmental sequence within a single stage; however, such intra-stage development is beyond the scope of the current study.) Given that an insufficient number of tasks may limit the number of emerged forms produced, leading to underestimation of development, two or three tasks were designed to elicit each structure to give sufficient opportunities to produce it. For each task, I provided an oral prompt and/or question with relevant visual materials, such as pictures and flyers shown on a computer monitor, along with the time limit for the response. The oral prompt and/or question was sometimes elab-orated or repeated when a participant did not understand fully. The pictures were

copyrighted images from Pixabay (https://pixabay.com), Openclipart (https://open clipart.org/), LibreShot (https://libreshot.com/), Flickr (https://www.flickr.com/), and Pexels (https://www.pexels.com/).

The specific tasks intended to produce the Stage 2 and Stage 3 structures were of several types. First, two tasks designed to elicit S neg V (e.g., *Students should not make a loud voice in a dormitory*) focused on making an announcement; and participants took the role of a movie theater manager in one task and a teacher in the other and were shown several pictures on which to base their production for each task. For the first task, they were asked to inform theater audiences of what they were or were not allowed to do during a movie, and, for the second, they were asked to explain a list of rules intended to make riding an elevator safe for students. In each task, participants were asked to produce one or two sentences relevant to each picture. Second, picture description tasks were designed to elicit Plural -s (e.g., *cookies*), Plural NP (e.g., *five bicycles*), and Progressive -*ing* (e.g., *Two players are jumping*). In these tasks, a series of pictures was given, accompanied by questions (*What are they? What are they doing?*). Each picture included different subjects and backgrounds. Third, tasks designed to elicit Possessive *'s* (e.g., *Jane's shoes*) required participants to describe objects in several pictures to answer various questions (*Whose is it? What color is it?*) and the directions to a particular building using a map that included several houses with owners' names on them. Finally, two narrative tasks designed to elicit Past -*ed* (e.g., *I enjoyed the time.*) required participants to describe past experiences (a winter holiday, first semester at university).

For the Stage 5 structure, three types of tasks were used to encourage participants to elicit Third person singular -*s* (e.g., *Internet provides useful information.*). In one task, participants were provided with flyers about a conference and asked to respond to a voice message based on the flyer. Two argumentative tasks required participants to present their opinions in response to particular questions (*Do you think that it is a good idea to live with a roommate? What do you think is the most useful way to prepare for a trip?*). For the final task, participants were asked to describe one of their family members.

Finally, two tasks were designed to elicit the Stage 6 structure, Cancel INV (e.g., *I wondered whether I could park my car*), both in the format of leaving a voice message, with the participant taking the role of a transfer student in one task, and as a potential guest of a hotel in the other. After reading through a script of a voice message from a university and a hotel flyer, respectively, the participants were required to leave a short voice message including at least three questions about their upcoming school life or their hotel reservation. These last two tasks did not intend to elicit particular types of questions; however, they were expected to give the participants opportunities to produce Cancel INV.

The time limits varied according to the required processing procedure and expected length of response and were based on time limits for similar tasks in standardized tests and a pilot study. The time limits were intended to elicit spontaneous responses by limiting the use of knowledge and favoring automatic processing in both modalities, following the assumptions of PT (Kawaguchi & Di Biase, 2012; Nicholas et al., 2019; Spinner, 2011), and to make the experimental settings as comparable as possible between the modalities. In particular, unlike interactive tasks in previous PT studies (e.g., interactions between participant and researcher or between two participants; Pienemann, 1998), where spontaneity is supported by the conversational format in which participants take turns, this study's speaking tasks were completed individually. Likewise, untimed writing may not favor spontaneous responses, as it allows writers as

much time as they like to monitor and revise their production as well as to use outside resources (see also Nicholas et al., 2019). In this regard, the time limits contributed to the establishment of similar experimental conditions in both modalities that encouraged the learners to complete the tasks spontaneously in an unplanned manner.

More specifically, some tasks, such as those designed to elicit S neg V and Possessive 's, required participants to respond spontaneously right after the oral prompt. In real-time conversation even in L1, a speaker typically takes a second (or two) to respond to an interlocutor's question and then starts to speak fluently (see also Butterworth, 1975; Lee, 2018). In accordance with this observation, Levelt's (1989) speaking model demonstrates that the processing procedure for oral production is incremental, progressing through different stages, which may take at least a brief moment to articulate. In this regard, this study considered a response within five seconds after the oral prompt to be a spontaneous response. Because the tasks asked participants to produce one or two sentences only, the time spent producing was not controlled unless it was obviously excessive. For the argumentative tasks in this study, the time limit was decided based on an argumentative task in the TOEFL Speaking component, which allows 15 seconds for planning and 45 seconds for production (one minute in total). However, in this study's pilot experiment using a one minute time limit, students at the low intermediate level did not fully complete the argumentative task, and the responses overall lacked contexts in which the Third person singular -s should be used. Therefore, the time limit for the argumentative tasks was extended to one and half minutes, considering the participants' varied proficiency levels. The longer time allowed meant participants were required to manage the assigned time limit of each task by themselves to plan and produce their own response; however, I sometimes intervened by asking further related questions when participants spent too much time planning (especially in spontaneous tasks) or when participants' responses were too short. Any production after the time limit was over was excluded from the dataset.

Writing tasks

The writing tasks were also conducted one-on-one. As with the speaking tasks, an oral prompt and/or question was given with relevant visual materials and the time limit, along with any additional guidance needed. The participant completed the tasks on a computer.[1] The oral prompt minimized the chances of the participants borrowing expressions from the prompts in their writing. The writing tasks were similar to the speaking tasks, with slight adjustments to more closely approximate real-life writing situations. Except for the modality, all aspects of the writing tasks were designed to be as comparable as possible to those of the speaking tasks. Because typing takes longer than speaking, the time limit for each writing task was twice that of the corresponding oral task.[2] Likewise, in spontaneous tasks, participants' responses were considered spontaneous if they started to type their response within five seconds.

---

[1]The participants were literate and fairly comfortable writing on the QWERTY keyboard.

[2]Because this study initially attempted to apply PT to L2 writing, it strived to follow general assumptions of PT and to make the experimental settings as comparable as possible across the oral and written modalities. The intention of doubling the time limits for the writing was to make the conditions comparable rather than to provide additional time for the monitoring and editing of writing. Hence, the length and structure of the participants' written task responses differed from what they might produce in a general writing task given more time. However, writing and speaking processes are distinctive regardless of time constraints (Polio, 2012; Williams, 2012), and the study's results indicate that the time constraints that were intended to create

For the Stage 2 and 3 features, similar to the equivalent oral tasks, several task types were employed. The two writing tasks designed to elicit S neg V required participants, first, to create a flyer to inform students of what they could and could not (or should not) do while living in a dormitory, and second, to write guidelines for a box of contact lenses, both based on pictures. Next, the participants were shown a series of pictures accompanied by questions to elicit Plural -s, Plural NP, and Progressive -ing. For Possessive 's, participants were asked to describe depicted objects, and to provide directions to someone's desk using a map of an office. Two narrative tasks required participants to describe the most memorable event they had attended and their last summer vacation with the intention of eliciting Past -ed.

For Stage 5, three types of tasks were designed to elicit Third person singular -s. In one task, participants responded to an email that included questions about a children's art exhibition based on information in a flyer. In two argumentative tasks, participants presented their opinions in response to particular questions (*Do you think it is a good idea to use a laptop during class? What do you think is the best way to find a job?*). In a description task, participants were prompted to describe one of their friends.

Finally, for the Stage 6 structure, there were two writing-an-email tasks in which the participants were supposed to act like new city residents and potential train passengers. After reading through an email from a residents' committee and a summary of a ticket, the participants were asked to write short request emails including at least three questions. Similar to the equivalent oral tasks, these two tasks were not designed to elicit particular types of questions. Rather, they left open the possibility of participants using Cancel INV (e.g., *I wondered whether I could book additional tickets for my children*).

In the speaking tasks, I sometimes intervened with additional explanation or encouragement to write more. As mentioned, the time limits for the speaking tasks were doubled for the corresponding writing tasks, an arbitrary increase intended to address the fact that typing requires more time than speaking (Table 2).

Procedure

The experiment was conducted in a quiet laboratory room in two sessions. I timed the participants' responses with an automatic stopwatch, and the participants could monitor the time limits while completing the tasks. The order of the speaking session and the writing session was counterbalanced; thus, half of the individuals completed the speaking tasks on the first day and the other half completed the writing tasks on the first day. The order of the tasks was pseudorandomized to separate those designed to elicit the same target structure by at least three other tasks. On the first day, the participants

---

similar levels of spontaneity in the two modalities did not entirely hinder processes specific to writing. That is, the written task responses showed evidence of diverse writing processes, including planning and revising. The exit questionnaire results confirm that the participants employed distinctive writing behaviors; for example: "*Although there is similar time limit in the writing task, I felt less pressure in the writing task because of the opportunity to edit while answering,*" "*I felt less anxious about the writing tasks, because I could revise ungrammatical structures,*" and "*Because I could check what I was writing, I tended to monitor and control my production more.*" In addition, the oral responses mostly had more words than the written responses (Table 3), which indirectly suggests that speaking and writing processes are quite different and that the participants took some time to check and edit their writing while answering. As mentioned in the conclusion, further research is needed to develop better measures of online writing behaviors.

Table 2. Summary of Speaking and Writing Tasks

Targeted structures	Speaking task		Writing task	
	Task type	Time limit	Task type	Time limit
S neg V	Making an announcement	Spontaneous	Creating a flyer	Spontaneous
	Making an announcement	Spontaneous	Writing product guidelines	Spontaneous
Plural –s, Plural NP, & Progressive –ing	Picture description task A	1 min	Picture description task A	2 min
	Picture description task B	1 min	Picture description task B	2 min
Possessive 's	Picture description task	Spontaneous	Picture description task	Spontaneous
	Giving directions via phone–call	Spontaneous	Giving directions via chatting	Spontaneous
Past –ed	Narrative task A	1.5 min	Narrative task A	3 min
	Narrative task B	1.5 min	Narrative task B	3 min
Third person singular –s	Response to a voice message	1.5 min	Response to an email	3 min
	Argumentative task A	1.5 min	Argumentative task A	3 min
	Argumentative task B	1.5 min	Argumentative task B	3 min
	Narrative task C	1.5 min	Narrative task C	3 min
Any types of questions (e.g., yes/no questions, *wh*–questions) & Cancel INV	Leaving a voice message A	1.5 min	Writing an email A	3 min
	Leaving a voice message B	1.5 min	Writing an email B	3 min

completed a background questionnaire about their demographic information; either the speaking or the writing tasks; and two working memory capacity tests (i.e., reading span test, operation span test) in counterbalanced order, as interruption tasks to mitigate any effects of their memory of the task types on their production on the second day. On the second day, they completed the tasks in the other mode and an exit questionnaire asking for their reflections on the experiment. They had a short break between each task, and additional short breaks if requested. Participants' answers in the speaking tasks were recorded with a Sony ICD-UX560F recorder, and Microsoft Word (with the spelling and grammar function blocked) was used for the writing tasks. The entire experiment took approximately two hours, and the participants received $30.

Measures

Emergence

According to PT, L2 development should be measured not by accuracy but by emergence, or "the point in time at which certain skills have, in principle, been attained or at which certain operations can, in principle, be carried out" (Pienemann, 1998, p. 138). The ability to use a structure systematically and productively, which shows the mapping of form and function in learners' processing (Pallotti, 2007), is not confirmed by a single token but by multiple tokens and/or multiple contexts of use. Therefore, following previous studies (Bonilla, 2015; Jansen, 2008; Pallotti, 2007; Pienemann, 1998; Spinner, 2011), the criterion of emergence for this study is arbitrarily set at four tokens and/or contexts of each morphosyntactic structure. Tokens occurring in formulaic language were excluded (Myles, 2004). For example, for syntax, if a participant

repeatedly used a hedging expression, such as *I do not think that…*, throughout the experiment, *I do not think* was not considered one of the contexts of S neg V; it was excluded as probable formulaic language. For morphology, a target item was considered to have emerged if it occurred with varied lexical words in at least four contexts and it was produced in morphological minimal pairs (e.g., *dog* vs. *dogs*), in creative constructions (i.e., overgeneralization, such as *goed* and *putted* instead of *went* and *put*), or with lexical variety (see Pallotti, 2007). Similarly, a repeatedly used target item in a chunk (e.g., *seems* in *it seems that…*) was not counted as one of the tokens of the Third person singular *-s*.

Accuracy

Following Lee and Spinner (2017), this study assesses the accuracy of morphosyntactic structures by means of obligatory occasion analysis (Ellis & Barkhuizen, 2005). All obligatory occasions for each structure were identified, and each occasion was then examined to investigate whether the correct structure was supplied. Then, the percentage of the correct forms was calculated. However, more than one use of identical forms in any single context was counted as a single use (e.g., *She like XX and sweet food. She like song…* counted as a single failure to supply an obligatory Third person singular *-s*).

This study defines the "accurate" use of a particular structure in three ways. First, when a structure is related to a verb in a sentence (e.g., Progressive *-ing*, Past *-ed*, and Third person singular *-s*), accuracy is determined in regard to tense, aspect, and subject-verb agreement (Ellis & Yuan, 2004). Second, the accuracy of other morphological structures (e.g., Plural NP) is coded as correct when the participants used the forms accurately and with correct agreement. Third, the accurate use of syntactic structures (e.g., S neg V) is determined by correct word order and inclusion of all required elements.

Coding and data analysis

Participants' oral responses were transcribed for the analysis. Grammatical and semantic errors were not corrected. In cases of self-correction of a target structure, only the last token was transcribed, and in cases of repetition, only one token was transcribed.

Two native speakers of English coded the participants' responses in the transcripts of both modalities (intercoder reliability = .77).[3] They counted the morphosyntactic structures predicted by PT and judged the accuracy of each usage. For all target structures produced more than four times, I double-checked the transcripts to ensure none of the usages were formulaic and to confirm the emergence of the structure (Pienemann, Johnston, & Brindley, 1988). While checking the coded data, I counted the number of cases of emergence in each participant's response and checked the stage in which each case of emergence occurred.

Following previous PT studies, this study created implicational tables for the participants' responses in each modality. Implicational scaling is a method of visually

---

[3]Intercoder reliability for the speaking tasks designed to elicit (1) S neg V = .82, (2) Plural *-s*, Plural NP, and Progressive *-ing* = .78, (3) Possessive '*s* = .75, (4) Past *-ed* = .76, and (5) Third person singular *-s* = . 80; for the writing tasks designed to elicit (1) S neg V = .77, (2) Plural *-s*, Plural NP, and Progressive *-ing* = .76, (3) Possessive '*s* = .80, (4) Past *-ed* = .82, and (5) Third person singular *-s* = . 74

representing L2 learners' developmental patterns in a table format. This arrangement of cross-sectional data allows for the examination of gradual progress over time, which possibly can be interpreted as representing L2 learning (Bonilla, 2015; Jansen, 2008; Pienemann, 1998). The first column of an implication table shows all participants arranged in order from less development in the top rows to greater development in the bottom rows. The rest of the columns indicate each stage predicted by PT from earlier stages (i.e., to the left). When the criteria for emergence were fulfilled in a learner's production, it is marked *e* (emergence). When the criteria for emergence were not fulfilled in a learner's production, it is marked N (none). In implicational scaling, a later stage structure that emerges earlier than an earlier stage structure is considered an error. Based on the number of errors, the predictability (coefficient of reproducibility; C of R) and reliability (coefficient of scalability; C of S) of the implicational scales were determined (Hatch & Lazaraton, 1991). C of R indicates the extent to which an implicational scaling predicts each participant's development based on his/her rank in the implicational scaling, and a C of R over 0.9 is considered to show acceptable predictability. The C of S is then calculated, with a C of S above 0.6 considered to confirm that an implicational scale is reliable. For accuracy, this study highlights emerged structures with accuracy over 0.75 (Lee & Spinner, 2017) in the implicational scales, which allows for investigating the extent to which the emerged structures are accurately used in learners' responses. Thus, the implicational tables for accuracy include target structures rather than the developmental stages predicated by PT in the columns. The order of the target structures in each stage in the columns was decided based on the number of cases of emergence across the participants' responses. The emerged structures that met the criterion of accuracy (i.e., 0.75) are in bold and highlighted.

To maintain consistency, the number of cases of emergence and the number of emerged structures whose accuracy is over .75 were considered to decide the order of the participants in the implicational tables. As discussed, participants with lower emergence numbers appear earlier in the tables, and participants with the same numbers are ordered based on ID numbers (from earlier to later), number of produced structures meeting the criterion of accuracy, and proficiency levels (from A2 to C1).

Next, a generalized mixed model was performed to understand whether emergence patterns differ between the oral and the written modalities, and a linear mixed model was performed to explore whether accuracy patterns differ between the modalities. Because the accuracy of the emerged structures in higher stages (e.g., Third person singular -*s* in stage 5) was usually lower than the criterion, a statistical analysis would have a large number of missing values, and the results might not be generalizable. Therefore, the linear mixed model was performed only with the five most frequent structures. Maximum likelihood estimation was used for both models.

Results

The data for analysis consisted of 2,816 oral responses from 87 participants and 2,752 written responses from 86 participants (the written responses of one participant were removed due to errors in the data). The average number of words in each proficiency level in tasks designed to elicit a particular structure is shown in Table 3. Overall, the participants produced more words in the speaking tasks ($\beta = -.14.12$, SE = 1.31, $t = -10.74$, $p < .001$); participants with higher proficiency also produced more

Table 3. Number of Words in Responses of Tasks Designed to Elicit Each Structure

	A2		B1		B2		C1	
	Oral	Written	Oral	Written	Oral	Written	Oral	Written
S neg V	53.21	38	69.30	50.71	68.91	51.60	89.61	53.50
	(17.27)	(9.20)	(22.24)	(9.95)	(19.15)	(13.94)	(43.14)	(6.64)
Plural –s,	27.11	18.35	36.18	26.31	35.39	27.89	46.39	33.83
Plural NP, &	(8.33)	(5.54)	(12.16)	(6.85)	(8.82)	(7.63)	(11.89)	(9.14)
Progressive								
–ing								
Possessive 's	45.07	46.04	54.09	57.52	60.21	56.74	62.50	59.93
	(10.94)	(11.12)	(16.79)	(14.58)	(19.70)	(11.80)	(19.04)	(16.52)
Past –ed	54.83	32.83	80.05	53.75	84.43	48.93	115.52	67.36
	(22.17)	(11.37)	(23.71)	(20.19)	(21.20)	(10.79)	(38.12)	(18.91)
Third–person	48.24	36.40	65.10	56	70.75	59.68	91.70	77.81
singular –s	(12.96)	(11.47)	(20.29)	(15.22)	(17.76)	(14.59)	(22.44)	(14.13)

Table 4. Generalized Mixed Model for Emergence

	Estimate	SE	z	p
Intercept	.79	.10	7.62	<. 001
Modality	−.21	.14	−1.49	.14

(e.g., C1: β = 28.35, SE = 3.46, t = 8.21, p <. 001) (Table 4).[4] Implicational scales were created using Microsoft Excel 2019, and other statistical analyses were computed, using the correlation and lmer4 package in R (R Core Team, 2018).

Emergence

PT posits that production of a specific structure in a stage confirms L2 learners' acquisition of the processing procedure in the stage. This study chose the most frequently emerged structure at Stage 2 (Plural -s) for the implicational scaling. Therefore, the target structures were a single word (Stage 1), Plural -s (Stage 2), Plural NP (Stage 3), Third person singular -s (Stage 5), and Cancel INV (Stage 6). Figure 1 shows the implicational scaling of oral and written responses. For the participants' oral responses, there were 435 available instances of emergence in the implicational scaling, and 1 error (C of R: 1; C of S: 0.99). The results confirmed that the implicational scaling successfully predicted the participants' L2 development in the oral modality with reliability. In particular, the structures in earlier stages emerged in most of the participants' production, whereas the structures in later stages emerged only in some advanced learners' production. For instance, emergence of structures in Stages 1 and 2, which require a lower level of processing procedures, occurred in all participants' responses, but emergence of structures in Stage 6, which require a higher level of processing procedures, was only found in five participants' responses.

For the participants' written responses, there were 430 available instances of emergence and 1 error in the implicational scaling (C of R: 1; C of S: 1). Similar to

---

[4]In the linear mixed model, number of words was the dependent variable; modality (oral, written) and proficiency (A2, B1, B2, C1) were fixed effects; and participant was a random effect. Oral and A2 were referential levels.
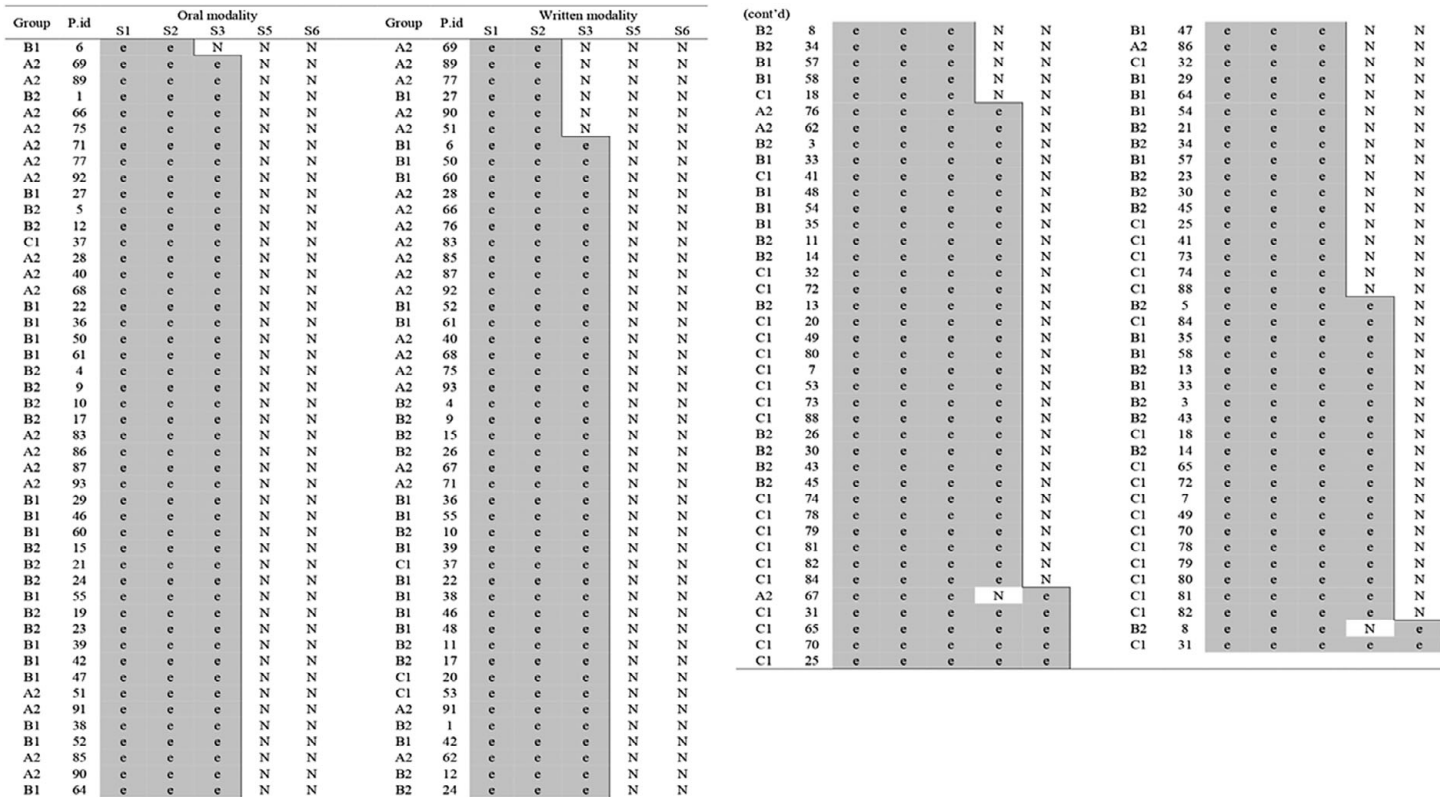
Group	P.id	Oral modality S1	S2	S3	S5	S6
B1	6	e	e	N	N	N
A2	69	e	e	e	N	N
A2	89	e	e	e	N	N
B2	1	e	e	e	N	N
A2	66	e	e	e	N	N
A2	75	e	e	e	N	N
A2	71	e	e	e	N	N
A2	77	e	e	e	N	N
A2	92	e	e	e	N	N
B1	27	e	e	e	N	N
B2	5	e	e	e	N	N
B2	12	e	e	e	N	N
C1	37	e	e	e	N	N
A2	28	e	e	e	N	N
A2	40	e	e	e	N	N
A2	68	e	e	e	N	N
B1	22	e	e	e	N	N
B1	36	e	e	e	N	N
B1	50	e	e	e	N	N
B1	61	e	e	e	N	N
B2	4	e	e	e	N	N
B2	9	e	e	e	N	N
B2	10	e	e	e	N	N
B2	17	e	e	e	N	N
A2	83	e	e	e	N	N
A2	86	e	e	e	N	N
A2	87	e	e	e	N	N
A2	93	e	e	e	N	N
B1	29	e	e	e	N	N
B1	46	e	e	e	N	N
B1	60	e	e	e	N	N
B2	15	e	e	e	N	N
B2	21	e	e	e	N	N
B2	24	e	e	e	N	N
B1	55	e	e	e	N	N
B2	19	e	e	e	N	N
B2	23	e	e	e	N	N
B1	39	e	e	e	N	N
B1	42	e	e	e	N	N
B1	47	e	e	e	N	N
A2	51	e	e	e	N	N
A2	91	e	e	e	N	N
B1	38	e	e	e	N	N
B1	52	e	e	e	N	N
A2	85	e	e	e	N	N
A2	90	e	e	e	N	N
B1	64	e	e	e	N	N

Group	P.id	Written modality S1	S2	S3	S5	S6
A2	69	e	e	N	N	N
A2	89	e	e	N	N	N
A2	77	e	e	N	N	N
B1	27	e	e	N	N	N
A2	90	e	e	N	N	N
A2	51	e	e	N	N	N
B1	6	e	e	e	N	N
B1	50	e	e	e	N	N
A2	28	e	e	e	N	N
A2	66	e	e	e	N	N
A2	76	e	e	e	N	N
A2	83	e	e	e	N	N
A2	85	e	e	e	N	N
A2	87	e	e	e	N	N
A2	92	e	e	e	N	N
B1	52	e	e	e	N	N
B1	61	e	e	e	N	N
A2	40	e	e	e	N	N
A2	68	e	e	e	N	N
A2	75	e	e	e	N	N
A2	93	e	e	e	N	N
B2	4	e	e	e	N	N
B2	9	e	e	e	N	N
B2	15	e	e	e	N	N
B2	26	e	e	e	N	N
A2	67	e	e	e	N	N
A2	71	e	e	e	N	N
B1	36	e	e	e	N	N
B1	55	e	e	e	N	N
B2	10	e	e	e	N	N
B1	39	e	e	e	N	N
C1	37	e	e	e	N	N
B1	22	e	e	e	N	N
B1	46	e	e	e	N	N
B1	48	e	e	e	N	N
B2	11	e	e	e	N	N
B2	17	e	e	e	N	N
C1	20	e	e	e	N	N
C1	53	e	e	e	N	N
A2	91	e	e	e	N	N
B2	1	e	e	e	N	N
B1	42	e	e	e	N	N
A2	62	e	e	e	N	N
B2	12	e	e	e	N	N
B2	24	e	e	e	N	N

(cont'd)

Group	P.id	S1	S2	S3	S5	S6
B2	8	e	e	e	N	N
B2	34	e	e	e	N	N
B1	57	e	e	e	N	N
B1	58	e	e	e	N	N
C1	18	e	e	e	N	N
A2	76	e	e	e	e	N
A2	62	e	e	e	e	N
B2	3	e	e	e	e	N
B1	33	e	e	e	e	N
C1	41	e	e	e	e	N
B1	48	e	e	e	e	N
B1	54	e	e	e	e	N
B1	35	e	e	e	e	N
B2	11	e	e	e	e	N
B2	14	e	e	e	e	N
C1	32	e	e	e	e	N
C1	72	e	e	e	e	N
B2	13	e	e	e	e	N
C1	20	e	e	e	e	N
C1	49	e	e	e	e	N
C1	80	e	e	e	e	N
C1	7	e	e	e	e	N
C1	53	e	e	e	e	N
C1	73	e	e	e	e	N
C1	88	e	e	e	e	N
B2	26	e	e	e	e	N
B2	30	e	e	e	e	N
B2	43	e	e	e	e	N
B2	45	e	e	e	e	N
C1	74	e	e	e	e	N
C1	78	e	e	e	e	N
C1	79	e	e	e	e	N
C1	81	e	e	e	e	N
C1	82	e	e	e	e	N
C1	84	e	e	e	e	N
A2	67	e	e	e	N	e
C1	31	e	e	e	e	e
C1	65	e	e	e	e	e
C1	70	e	e	e	e	e
C1	25	e	e	e	e	e

Group	P.id	S1	S2	S3	S5	S6
B1	47	e	e	e	N	N
A2	86	e	e	e	N	N
C1	32	e	e	e	N	N
B1	29	e	e	e	N	N
B1	64	e	e	e	N	N
B1	54	e	e	e	N	N
B2	21	e	e	e	N	N
B2	34	e	e	e	N	N
B1	57	e	e	e	N	N
B2	23	e	e	e	N	N
B2	30	e	e	e	N	N
B2	45	e	e	e	N	N
C1	25	e	e	e	N	N
C1	41	e	e	e	N	N
C1	73	e	e	e	N	N
C1	74	e	e	e	N	N
C1	88	e	e	e	N	N
B2	5	e	e	e	e	N
C1	84	e	e	e	e	N
B1	35	e	e	e	e	N
B1	58	e	e	e	e	N
B2	13	e	e	e	e	N
B1	33	e	e	e	e	N
B2	3	e	e	e	e	N
B2	43	e	e	e	e	N
C1	18	e	e	e	e	N
B2	14	e	e	e	e	N
C1	65	e	e	e	e	N
C1	72	e	e	e	e	N
C1	7	e	e	e	e	N
C1	49	e	e	e	e	N
C1	70	e	e	e	e	N
C1	78	e	e	e	e	N
C1	79	e	e	e	e	N
C1	80	e	e	e	e	N
C1	81	e	e	e	e	N
C1	82	e	e	e	e	N
B2	8	e	e	e	N	e
C1	31	e	e	e	e	e

Figure 1. Implicational scaling of the oral and written responses.
Note: P.id = participant's assigned id #; S in S1, S2, S3, S5, S6 = stage.

the results in the oral modality, these results indicate that the order of the participants' L2 development in their written production could be described accurately and reliably by PT. The emergence of the target structures in the participants' written production was similar to that in their oral production. The structures emerged gradually and in the predicted order. However, in some cases, structures emerged later in the written modality. For instance, a higher percentage of participants failed to show acquisition of the processing procedures in Stage 3 in their written responses than in their oral responses (6.98% in the written modality vs. 1.14% in the oral modality). Likewise, a lower percentage of participants showed the emergence of higher stage structures in the written modality than in the oral modality; specifically, the Third person singular (Stage 5; 24.41% in the written modality vs. 39.08% in the oral modality) and Cancel INV (Stage 6; 2.32% in the written modality vs. 5.74% in the oral modality).

Overall, the emergence patterns in the two modalities were similar. A greater number of participants showed emergence in earlier stages, with decreasing numbers of participants showing emergence of structures at later stages. In addition, the results of a majority of participants were consistent between the oral and written modalities ($N = 56$); in other words, emergence occurred at the same stage across the modalities. Specifically, participants in the A2 group reached Stage 3 in both modalities, but some participants reached Stage 5 starting at the B1 level (B1: 11.76%; B2: 26.67%; C1: 81.81%) and some of the C1 group even reached Stage 6 (9.09%), again in both modalities. However, some individual participants showed the emergence of more structures in the oral modality than in the written modality ($N = 25$), and vice versa ($N = 5$), although the number of the participants who showed consistent results between the oral and written modalities was statistically greater than that of the participants who showed inconsistent results [χ^2 (1, $N = 86$) = 7.86, $p = .005$].

To explore the similarity and difference between the implicational scales for the two modalities, Pearson correlations were conducted with the number of participants who had reached each stage (Figure 2). For instance, 81 (out of 86) participants showed the
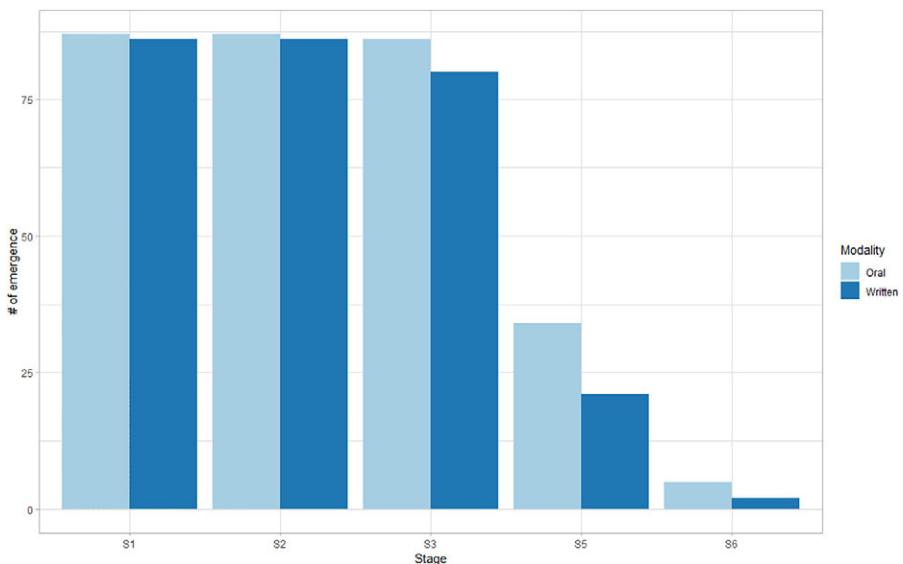
Figure 2. Number of participants whose production showed emergence at each stage.

Table 5. Percentage of Emergence and Cs of R for Individual Structures in Both Modalities

Stage	Structure	Oral modality		Written modality	
		% of emergence	C of R	% of emergence	C of R
Stage 1	Words	100	1	100	1
Stage 2	S neg V	97.70	0.98	95.35	0.97
	Plural –s	100	1	100	1
	Progressive –ing	100	1	98.83	0.99
	Possessive 's	89.66	0.91	90.70	0.95
	Past –ed	48.28	0.48	45.34	0.48
Stage 3	Plural NP	98.85	1	93.02	0.95
Stage 5	Third person singular –s	39.10	0.99	24.41	0.99
Stage 6	Cancel INV	5.75	0.99	2.33	1

emergence of the Stage 3 structures in their written responses; thus, the number of participants counted as having reached Stage 3 was 81. The results of the correlation analysis indicated that the implicational scales for the two modalities were correlated for all participants ($r = .99$, $p = .001$).

The differences between the two modalities were not significant; however, at all proficiency levels there was slightly higher emergence in the oral modality than the written modality at Stages 3, 5, and 6, with the greatest cross-modality difference in the production of the Stage 5 structure.

A generalized mixed model was performed with emergence of a structure (emergence, no emergence) as a dependent variable, modality (oral, written) as a fixed effect, and participant as a random effect. Oral modality served as the reference level. Modality was not significant in this model ($\beta = -.21$, SE = .14, $z = -1.49$, $p = .14$), indicating that the pattern of the gradual emergence of the target structures was similar in the two modalities.

Furthermore, the ratio of emergence and a set of Cs of R were calculated for individual structures in the two implicational tables (Table 5). The ratio of emergence and the Cs of R for individual structures indicate the extent to which each element occurs similarly or differently between the modalities and whether predictability of a particular structure being produced differs depending on modality. In other words, these analyses had the potential to reveal subtle similarities and differences between the modalities. The ratio of emergence and the Cs of R seem comparable between the oral and written modalities. In particular, although the percentage of emergence was slightly higher in the oral modality in Stage 2, overall, the same order of emergence appeared in both modalities: Plural -s, Progressive -ing, S neg V, Possessive 's, Past -ed. In addition, in both modalities, the Cs of R for all structures except for Past -ed were over. 9. Past -ed showed a lower C of R than the criterion, suggesting that reaching a certain developmental stage does not necessarily mean that all possible structures in this stage emerge, which is consistent with previous PT studies (e.g., Di Biase et al., 2015).

Accuracy

Following Lee and Spinner (2017), accuracy below. 75 is considered to indicate that a learner cannot consistently produce the structure in question accurately. Figure 3 illustrates the accuracy of the participants' oral production together with emergence. The gray cells indicate emerged structures, with the number in bold for accuracy over. 75. Empty cells show that a structure has not emerged (thus accuracy cannot be

Figure 3. Accuracy of the oral and written responses.
Note: C = Cancel INV; P = Plural NP; P. id = participant's assigned id #; S in S2, S3, S5, S6 = stage.

Table 6. Summary of Accuracy

Structure	# (%) of participants who showed emergence of the structures and met the criterion of accuracy	
	Oral modality	Written modality
S neg V (e.g., **You can't take** photos of the movies; *<u>don't smoking</u>.*)	66 (.78)	67 (.82)
Plural –s (e.g., Brown **shoes**; *some <u>fire–fighter</u>)	55 (.63)	68 (.79)
Progressive –ing (e.g., They **are playing** the basketball; *he <u>dribbling</u>; *he <u>dribble</u>)	42 (.48)	54 (.64)
Possessive 's (e.g., Across the deli shop, there is **Lily's** house; *<u>Oliver desk</u> is in front of <u>Emma desk</u>.)	57 (.73)	58 (.75)
Past –ed (e.g., I **watched** some movie; *we just <u>studying and play</u> the game.)	25 (.60)	30 (.77)
Plural NP (e.g., **Two players**; *<u>five bicycle</u>)	29 (.34)	54 (.68)
Third–person singular –s (e.g., Internet **provides** lots of information; *you <u>feels</u> very uncomfortable.; *he <u>like</u> to play soccer.)	16 (.46)	17 (.81)
Cancel INV (e.g., **I'm just wondering how much the breakfast is**; *<u>I wonder that breakfast is included?</u>)	2 (.40)	0

Note: Accurate forms are bold; inaccurate forms are underlined.

calculated). Zero indicates consistently inaccurate production. The analysis considered only the accuracy of emerged responses.

The oral responses showed emergence of the early-stage structures (e.g., Plural -s and Progressive -ing) for most participants, but accuracy varied. Overall, the participants who showed greater development produced more of the early-stage structures more accurately. The number of emerged structures that met the criterion of accuracy is limited in the top rows (e.g., $N = 0$ for participant 69) and increases steadily in the bottom rows (e.g., $N = 5$ for participant 25). In other words, participants' responses showed a gradually increasing ability to use these structures accurately. In addition, compared to the structures in the later stages, the participants produced the structures in the earlier stages more accurately. For instance, both S neg V (Stage 2) and plural NP (Stage 3) emerged in most of the participants' responses ($N = 85$ and $N = 86$, respectively). However, 69 participants (78%) produced S neg V with accuracy over. 75, whereas only 31 participants (34%) reached the criterion in the production of Plural NP (34%).

In the written responses as well, accuracy in the production of a particular structure increased along with the participants' L2 development. However, compared to their oral responses, in their written responses the percentage of the emerged responses that met the criterion of accuracy was relatively stable, regardless of the difficulty of the structures; in other words, the percentages were similar for structures in earlier and later stages (Table 6; Figure 3).

To explore differences in general trends of accuracy between the oral and written modalities, the next analysis considered only the accuracy of the five most frequently emerged structures. In a linear mixed model, accuracy was the dependent variable; modality (oral, written) and structure (S neg V, Plural -s, Progressive -ing, Possessive 's, Plural NP) were fixed effects; participant was a random effect. The oral modality and Plural NP served as reference groups. Depending on type of structure, accuracy differed. The difference in accuracy of each structure seemed aligned with the order

Table 7. Linear Mixed Model for Accuracy

Intercept	Estimate	SE	t	p
Modality	.07	.01	7.08	<. 001
S neg V	.15	.01	10.35	<. 001
Plural –s	.09	.01	5.89	<. 001
Progressive –ing	−.01	.01	−.83	.41
Possessive 's	.10	.02	6.93	<. 001

of the designated columns for the structures in the implicational table. In particular, Plural NP, which appears in the rightmost of the columns for the five structures in the implicational table, revealed lower accuracy than S neg V (β = .15, SE = .01, t = 10.35, p <. 001), Plural -s (β = .09, SE = .01, t = 5.89, p <. 001), and Possessive 's (β = .10, SE = .02, t = 6.93, p <. 001). However, there was no significant difference in the accuracy of Plural NP or of Progressive -*ing* (β = −.01, SE = .01, t = −.83, p = .41). More importantly, the participants produced the target structures more accurately in the written mode than in the oral mode (β = .07, SE = .01, t = 7.08, p <. 001) (Table 7).

Discussion

This study assesses the applicability of PT to L2 writing by comparing how well PT predicts development in oral and written modalities. In addition, the study compares the accuracy of the morphosyntactic structures that correspond to PT's stages in learners' oral and written production. A series of communicative tasks required participants to produce spontaneous responses in oral and written modalities. The results of this cross-sectional study imply that the structures in the earlier stages of PT are likely to be produced earlier than the structures in the later stages and that learners become more and more able to produce the structures in later stages as their L2 develops. With a few negligible exceptions, the emergence of structures showed developmental patterns to be comparable between the oral and written modalities. Furthermore, accuracy seemed to be aligned with emergence: Structures that emerged earlier were more likely to be produced with higher accuracy than structures that emerged later. In addition, the results suggest that particular structures are not necessarily produced with accuracy when they first emerge. The learners showed a gradual increase in accuracy as their L2 development progressed. This gradual development in accuracy was found in both oral and written modalities but accuracy was higher, and increased more rapidly, in the written modality than in the oral modality.

Emergence

Although each modality has unique characteristics (e.g., Gilabert et al., 2016; Williams, 2012), this study demonstrates that a single framework can be used to assess L2 learners' language development in their written production as well as their spoken production. While some differences between oral and written modalities were observed, the overlap between the two modalities led to comparable developmental patterns. Specifically, the participants revised and edited their responses more freely in the writing tasks than in the speaking tasks (Byrnes & Manchón, 2014; Granfeldt, 2008). Although the participants also corrected some of their responses in the speaking tasks, this behavior was limited to when they noticed having made a grammatical error right after uttering it. In

this case, the learners usually self-repaired, such as in *the man ride riding is riding*. It is often assumed that L2 learners more frequently avoid producing structures at higher levels when speaking than when writing (Kuiken & Vedder, 2011; Son, 2022; Tavakoli, 2014), but this study's participants appeared neither to avoid producing higher level structures in the speaking tasks nor to attempt to produce more higher level structures (over the criterion of emergence) in the writing tasks. Thus, some differences between oral and written modalities did not affect the overall developmental patterns.

The study sought to address some limitations in the few theory-based previous studies of L2 language development in writing (Boss, 2008; Byrnes & Sinicrope, 2009; Weissberg, 2006). Overall, the previous studies left open the possibility of extending one of the theoretical frameworks of L2 development based on oral production to written production, although they also reported some inconsistencies with the predictions of the frameworks, which may be due to these studies' use of (a) insufficient numbers of tasks designed to elicit particular structures, (b) non-communicative tasks, and/or (c) inappropriately designed tasks for the purpose of eliciting particular structures. In addition, no previous study has sought to compare the application of a certain theoretical framework of L2 development in oral and written modalities.

Hence, to complement and extend the aforementioned research, this study designed a series of communicative tasks, each intended to elicit one of the morphosyntactic structures predicted by PT to align with the developmental stages. With some slight modifications, identical task types were used in the two modalities. With this methodological improvement, this study revealed similar developmental patterns between the oral modality and the written modality and provided evidence for L2 language development in the written modality. The study's application of PT's theoretical framework to both oral and written modalities is an important leap in the research on language development.

This study's demonstration of the possibility of using PT to investigate L2 learners' development in the written modality also supports Pienemann's (1998) steadiness hypothesis, which posits that, in oral production, L2 learners are able to process morphosyntactic structures consistently regardless of type of communicative task, if the tasks are designed to use the same knowledge and skills. The current study's use of comparable tasks suggests that such consistency in oral production might extend to written production. It is assumed in this study that the best way to observe parallels between oral and written production is to use tasks that encourage the use of similar types of knowledge and skills. However, in some cases L2 learners may be able to deploy explicit knowledge and metalinguistic L2 knowledge in writing but not in speaking, due to the possibility of revision in writing (e.g., Schoonen et al., 2009; Williams, 2012). The time constraints in this study were intended to reduce the possibility of the learners using explicit knowledge in their written production, encouraging spontaneous production and leading them to rely more on L2 implicit knowledge in both modalities.

The use of comparable tasks and time constraints encouraged L2 learners to employ the same knowledge in both modalities, in keeping with the steadiness hypothesis, which I argue can now be extended to certain writing tasks. Whether the developmental order of PT is found in other types of writing tasks, such as essays with multiple revisions, has yet to be investigated, which calls for further research. In particular, this study doubled the time limits of the speaking tasks to set time limits for the writing tasks in order to make the experimental settings comparable. However, if future research can determine longer time constraints that still maintain the spontaneity of written

responses, the allowance of additional time to revise might show more emergence of structures at a higher level in written production.

Håkansson and Norrby (2007) also argued that L2 learners followed the developmental order predicted by PT in their writing. However, the current study's findings provide clearer evidence, for a few reasons. First, Håkansson and Norrby tested only advanced learners, while this study included a wide range of proficiency levels, enabling it to observe that learners followed the developmental order predicted by PT. This finding means that the applicability of PT's prediction to L2 written production can be more widely generalized. In addition, Håkansson and Norrby used translation tasks to assess writing, while this study utilized similar communicative speaking and writing tasks, so that, in both modalities, the participants had to come up with their own ideas and produce morphosyntactic structures by themselves within the designated time.

Accuracy

The development of accuracy may be relevant to PT's perspective on automatization (Pienemann, 1998). L2 learners may need some time for underlying processing routines to become automatized; even after a particular morphosyntactic structure emerges in L2 learners' production, its accuracy is not guaranteed. This study's results showed that L2 learners were able to produce morphosyntactic structures more accurately along with their further L2 development, as shown in the emergence of structures, and that their accuracy with structures that appeared earlier was higher than their accuracy with structures that appeared later. These results were similar between the two modalities.

However, the L2 learners reached the criterion of accuracy and showed greater stability of accuracy, much earlier in their written production than in their oral production. These results indicate that while the different characteristics of the oral and written modalities may not have changed the overall pattern of emergence (Research Question 1), they may lead to differences in accuracy (Research Question 2). Writing, even with time constraints, allows learners opportunities to plan, check, and edit grammatical errors in their production (Kellogg, 1996, 2001; McCutchen, 1996; Och, 1979). In addition, the written modality both allows and encourages L2 learners to use metalinguistic knowledge to monitor their language and accuracy (Williams, 2012), which may lead to greater accuracy compared to speaking, in which they rely on implicit knowledge (Polio, 2012). Furthermore, as Schoonen, Snellings, Stevenson, & Van Gelderen (2009) argued, spoken language is more tolerant of errors than written language, which may lead L2 learners to monitor their language more thoroughly in written tasks than in oral tasks. Hence, the unique characteristics of the written modality may have led this study's participants to pay more attention to the structures they produced and to correct more grammatical errors in the written tasks than in the oral tasks.

The differences in accuracy between oral and written production are also partially aligned with some previous studies that demonstrated the effects of task modality on L2 learners' production (e.g., Kormos, 2014; Kormos & Trebits, 2009; Tavakoli, 2014; Zalbidea, 2017). Although the current study employed different measures for accuracy (i.e., means of obligatory occasion analysis vs. error per T-unit; see also, Polio & Shea, 2014), its findings are consistent with those of the previous studies and provide more evidence for the effects of the written modality's characteristics that can lead to greater accuracy.

Conclusion

This study sheds light on the possibility of using the same theory to investigate L2 language development in the oral modality and the written modality. The study used similar tasks in oral and written modalities and compared the results to provide a better understanding of the applicability of a language development theory, PT, to writing. The results suggest that L2 linguistic knowledge may be developed in similar ways regardless of modality and demonstrate that theory-based investigations of language development are possible in the written modality as well as in the oral modality.

In addition, the study assessed whether accuracy is aligned with the emergence of morphosyntactic structures. The results showed a gradual progress in accuracy, with patterns similar to those of emergence, although the written modality exhibited a more consistent progression, possibly due to the opportunities to revise and edit writing.

This study has some limitations, which point to directions for further research. First, the time constraints for both task types were intended to encourage the learners to respond spontaneously. However, previous PT studies have not provided a principled way of determining appropriate time constraints for written tasks. Hence, the time constraint for each written task was set, rather arbitrarily, at double the time for the task's oral counterpart, considering the different speeds of typing and speaking. It is possible that the different time constraints between the modalities unintentionally led to different degrees of spontaneity, specifically, less spontaneity in the learners' written production. Further research should investigate the optimal time constraints for eliciting spontaneous production in both modalities as well as whether PT can be extended to L2 writing without time constraints.

Next, this study did not analyze L2 learners' online writing behaviors, but doing so could provide concrete evidence for the effects of the different characteristics of the two modalities on accuracy. The study assumed that the unique characteristics of writing affected accuracy. However, measuring online writing behaviors with the use of keystroke logging (e.g., Leijten & Van Waes, 2013), stimulated recall (e.g., De Silva & Graham, 2015), or both (e.g., Révész, Kourtali, & Mazgutova, 2017) would provide information on whether and how L2 learners actually do revise and edit their grammatical errors while writing, and whether such behaviors do lead to higher accuracy in their writing than in their speaking.

Finally, this study followed Lee and Spinner (2017) in setting 75% as the criterion of accuracy. Because previous PT studies did not explore the accuracy and emergence of morphosyntactic structures together, there is little evidence for whether this criterion is valid. Thus, the validity of the criterion of accuracy is another issue for future research to address.

References

Anderson, J. (1992). Automaticity and the ACT theory. *The American Journal of Psychology*, *105*, 165–180.

Bonilla, C. (2015). From number agreement to the subjunctive: Evidence for processability theory in L2 Spanish. *Second Language Research*, *31*, 53–74.

Boss, B. (2008). Exploring the acquisition of German verb morphology by instructed learners. *Australian Review of Applied Linguistics*, *31*, 14.1–14.13.

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology*, *29*, 591–620.

Bourdin, B., & Fayol, M. (2000). Is graphic activity cognitively costly? A developmental approach. *Reading and Writing*, *13*, 183–196.

Brown, J., McDonald, J., Brown, T., & Carr, T. (1988). Adapting to processing demands in discourse production: The case of handwriting. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 45–59.

Bulté, B., & Housen, A. (2009, September 13–16). *The development of lexical proficiency in L2 speaking and writing tasks by Dutch-speaking learners of French in Brussels* [Paper presentation]. Task-Based Language Teaching Conference, Lancaster, England.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, *4*, 75–87.

Byrnes H., & Manchón, R. (2014). Task-based language learning: Insights from and for L2 writing: An introduction. In H. Byrnes & R. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 1–27). John Benjamins.

Byrnes, H., & Sinicrope, C. (2009). Advancedness and the development of relativization in L2 German: A curriculum-based longitudinal study. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 125–154). Routledge.

Cumming, A. (2012). Goal theory and second-language writing development, two ways. In R. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 165–190). De Gruyter Mouton.

De Silva, R., & Graham, S. (2015). The effects of strategy instruction on writing strategy use for students of different proficiency levels. *System*, *53*, 47–59.

Di Biase, B., Kawaguchi, S., & Yamaguchi, Y. (2015). The development of English as a second language. In C. Bettoni & B. Di Biase (Eds.), *grammatical development in second languages: Exploring the boundaries of processability theory* (pp. 85–115). European Second Language Association.

Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics*, *18*, 4–22.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford University Press.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*, 59–84.

Gass, S. (1980). An investigation of syntactic transfer in adult second language learners. In R. C. Scarcella & S. D. Krashen (Eds.), *Research in second language acquisition: Selected papers of the Second Language Acquisition Research Forum: Issues in Second Language Research* (pp. 132–141). Newbury House.

Gilabert, R., Manchón, R., & Vasylets, O. (2016). Mode in theoretical and empirical TBLT research: Advancing research agendas. *Annual Review of Applied Linguistics*, *36*, 117–135.

Granfeldt, J. (2008). Speaking and writing in French L2: Exploring effects on fluency, complexity and accuracy. In S. van Daele, A. Housen, F. Kuiken, M. Pierrard, & I. Vedder (Eds.), *Complexity, accuracy and fluency in second language use, learning and teaching* (pp. 87–98). University of Brussels.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguists*. Heinle & Heinle.

Håkansson, G., & Norrby, G. (2007). Processability theory applied to written and oral Swedish. In F. Mansouri (Ed.), *Second language acquisition research: Theory-construction and testing* (pp. 81–94). Cambridge Scholars.

Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us?. *Applied Psycholinguistics*, *30*, 555–582.

Jansen, L. (2008). Acquisition of German word order in tutored learners: A cross-sectional study in a wider theoretical context. *Language Learning*, *58*, 185–231.

Kaplan, R., & Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 173–281). MIT Press.

Kawaguchi, S., & Di Biase, B. (2012). Acquiring procedural skills in L2: Processability theory and skill acquisition. *Studies in Language Sciences*, *11*, 68–95.

Keenan, E., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, *8*, 63–99.

Kellogg, R. (1996). A model of working memory in writing. In C. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–71). Lawrence Erlbaum.

Kellogg, R. (2001). Competition for working memory among writing processes. *American Journal of Psychology*, *114*, 175.

Kormos, J. (2014). Differences across modalities of performance. In H. Byrnes & R. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 193–216). John Benjamins.

Kormos, J., & Trebits, A. (2009, September 13–16). *Task-related differences across modes of performance* [Paper presentation]. Task-Based Language Teaching Conference, Lancaster, England.

Kuiken, F., & Vedder, I. (2011). Task complexity and linguistic performance in L2 writing and speaking. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 91–104). John Benjamins.

Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, *43*, 781–812.

Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, *12*, 439–448.

Lee, S. (2018). *Effective planning in real-time speaking test tasks* [Unpublished doctoral dissertation]. Michigan State University.

Lee, S., & Spinner, P. (2017, October 12–15). *The place of accuracy in processability theory* [Paper presentation]. Second Language Research Forum, Columbus, OH.

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*, 358–392.

Levelt, W. (1989). *Speaking: From intention to articulation*. MIT Press.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, *8*, 299–325.

Myles, F. (2004). From data to theory: The overrepresentation of linguistic knowledge in SLA. *Transactions of the Philological Society*, *102*, 139–168.

Nicholas, H., Lenzing, A., & Ross, J. (2019). How does PT's view of acquisition relate to the challenge of widening perspectives on SLA? In A. Lenzing, H. Nicholas, & J. Ross (Eds.), *Widening contexts for Processability Theory: Theories and issues* (pp. 391–398). John Benjamins.

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 717–761). Blackwell.

Ochs, E. (1979). Planned and unplanned discourse. In T. Givón (Ed.), *Discourse and syntax* (pp. 51–80). Brill.

Pallotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics*, *2*, 361–382.

Papageorgiou, S., Tannenbaum, R., Bridgeman, B., & Cho, Y. (2015). The association between TOEFL iBT® test scores and the Common European Framework of Reference (CEFR) levels [Research Memorandum No. RM-15-06]. Educational Testing Service.

Park, J. (2017). *Syntactic complexity as a predictor of second language writing proficiency and writing quality* [Unpublished doctoral dissertation]. Michigan State University.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. John Benjamins.

Pienemann, M. (2005). An introduction to processability theory. In M. Pienemann (Ed.), *Cross-linguistic aspects of processability theory* (pp. 1–60). John Benjamins.

Pienemann, M., Di Biase, B., Kawaguchi, S., & Håkansson, G. (2005). Processing constraints on L1 transfer. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 128–153). Oxford University Press.

Pienemann, M., & Håkansson, G. (1999). A unified approach toward the development of Swedish as L2: A processability account. *Studies in Second Language Acquisition*, *21*, 383–420.

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition*, *10*, 217–243.

Pienemann, M., & Lenzing, A. (2020). Processability Theory 1. In B. VanPatten, G. Keating, & S. Wulff (Eds.), *Theories in second language acquisition* (pp. 162–191). Routledge.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, *21*, 375–389.

Polio, C. (2017). Second language writing development: A research agenda. *Language Teaching*, *50*, 261–275.

Polio, C., & Park, J. (2016). Language development in second language writing. In R. Manchón & P. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 287–306). De Gruyter Mouton.

Polio, C., & Shea, M. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10–27.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language*, *29*, 417–447.

Révész, A., Kourtali, N. E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, *67*, 208–241.

Sasaki, M. (2004). A multiple-data analysis of the 3.5 year development of EFL student writers. *Language Learning*, *54*, 525–582.

Schoonen, R., Snellings, P., Stevenson, M., & Van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 77–101). Multilingual Matters.

Serrano, R., Tragant, E., & Llanes, À. (2012). A longitudinal analysis of the effects of one year abroad. *Canadian Modern Language Review*, *68*, 138–163.

Son, M. (2022). The role of modality and working memory capacity in L2 production. *Language Teaching Research*.

Spinner, P. (2011). Second language assessment and morphosyntactic development. *Studies in Second Language Acquisition*, *33*, 529–561.

Tardy, C. (2012). A rhetorical genre theory perspective on L2 writing development. In R. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 165–190). De Gruyter Mouton.

Tavakoli, P. (2014). Storyline complexity and syntactic complexity in writing and speaking tasks. In H. Byrnes & R. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 217–236). John Benjamins.

Vasylets, O., Gilabert, R., & Manchón, R. M. (2017). The effects of mode and task complexity on second language production. *Language Learning*, *67*, 394–430.

Weissberg, B. (2000). Developmental relationships in the acquisition of English syntax: Writing vs. speech. *Learning and Instruction*, *10*, 37–53.

Weissberg, R. (2006). *Connecting speaking & writing in second language writing instruction*. University of Michigan Press.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, *21*, 321–331.

Zalbidea, J. (2017). "One task fits all"? The roles of task complexity, modality, and working memory capacity in L2 performance. *The Modern Language Journal*, *101*, 335–352.

Zalbidea, J. (2021). On the scope of output in SLA: Task modality, salience, L2 grammar noticing, and development. *Studies in Second Language Acquisition*, *43*, 50–82.