＜論文＞

Development of clausal and subclausal grammatical complexity and their relationship to overall length in second language writing over a year

Daisuke NAKAMURA

Abstract

The present study investigates the development of the overall length, clausal (coordination and subordination) and subclausal grammatical complexity of paragraphs written by second language learners, employing both general indices for these dimensions and fine-grained measures of nominal complexity included in automated text analysing tools of grammatical complexity, the L2SCA and the TAASSC. Adult learners of English with an intermediate level of proficiency (N＝17) wrote paragraphs on eight topics over a year. The results showed that : a) grammatical complexity developed in order from subordination through phrasal elaboration and overall length to coordination, b) learners made a sentence longer by adding more subordinate clauses in the initial phase and by lengthening a clause in the final phases of development, and c) nominal subtypes also showed diverse developmental speeds with commonalities in that modifying elements were initially and finally added onto nominals in general and the nominals in subject and direct object slots in particular, with their reductions in the middle phase. The paper concluded that, if general developmental trajectories from coordination through subordination to phrasal elaboration found in the literature were interpreted in terms of overall length rather than via the time of learning, the trajectories were robust.

Keywords : *grammatical complexity, developmental trajectories, clausal and subclausal complexity indices, second language writing, longitudinal studies*

Introduction

Constructing complex sentences is clearly one hallmark of advanced second language (L2) writing proficiency ; most textbooks of academic writing (e.g., Hogue, 2008) have dimensions of grammatical complexity in their rubric. In previous studies of written L2, grammatical complexity was operationalised as lengthening units of analyses, such as a sentence or a terminable unit (T-unit), a main clause plus any dependent clauses attached to it (Hunt, 1965). In general, as writers' L2 proficiency increases, so does the length of their essays (Wolfe-Quintero, Inagaki, & Kim, 1998), as confirmed by Kern and Schultz (1992) and Larsen-Freeman (1983). Nonetheless, grammatical complexity is not a monolithic concept ; for example, sentential, clausal, and phrasal complexity should be distinguished as subdimensions of syntactic complexity (Bulte & Housen, 2014 ; Norris & Ortega, 2009). As a result, studies employing multiple measures of different dimensions, including traditional length-based ones, are necessary in order to capture detailed tracking of the development of L2 grammatical complexity (Norris & Ortega, 2009). In addition, recent research has begun to employ more specific indices of grammatical complexity such as the number of words before main verbs (Crossley & McNamara, 2014) and that of dependents per nominal subject (Kyle, 2016). These 'finer-grained' measures can capture developmental differences in a particular subdimension (such as phrasal complexity) that traditional, 'coarse-grained' measures such as the mean length of clause (MLC) cannot reveal.

Thus, the present paper aims to consider the development of grammatical complexity in paragraphs written by L2 learners with an intermediate level of proficiency in terms of overall length-based complexity, and clausal and phrasal complexity over a year, employing both traditional 'coarse-

日本女子体育大学（講師）

grained' and newly-developed 'fine-grained' indices of grammatical complexity, using automated text analysis tools provided by Lu (2010) and Kyle (2016).

Development of overall length, clausal and phrasal complexity in L2 writing

Compared to cross-sectional studies, the number of the longitudinal studies is somewhat small in second language acquisition (SLA) literature in general (Ortega & Byrnes, 2008a) and in L2 writing research in particular (Ortega, 2003), even though such studies are indispensable for clarifying advanced L2 capacities (Ortega & Byrnes, 2008b). Only six longitudinal L2 writing research studies, compared to 21 cross-sectional ones, were identified in Ortega's (2003) research synthesis. These earlier studies employed length-based measures such as the mean length of T-units (MLT, see Hunt, 1965) as complexity indices, and showed that, as the learners' proficiency increased or course proceeded, so did the length of their writing (Casanave, 1994 ; Kern & Shultz, 1992 ; Larsen-Freeman, 1983, 2006 ; see Arthur, 1979, for the null development). Kern and Shultz (1992), for example, investigated L2 essays written by L2 learners of French over eight months. Learners were asked to write essays on four different topics during class and their grammatical complexity was measured via MLT. The results showed the pattern of initial moderate increases through sudden rises to final drops in MLT. Similar increases in MLT were found in Larsen-Freeman's (1983) study of an EFL population.

However, relying on length-based measures exclusively cannot capture developmental intricacies in grammatical complexity neatly since, as Norris and Ortega (2009) pointed out, learners can make sentences longer by complexifying clauses (such as adding subordinations) and/or phrases (such as adding prepositional phrases). Since overall trends from coordination through subordination to phrasal elaboration in the development of grammatical complexity were found in the literature, Norris and Ortega (2009) claimed the necessity of complexity indices for these dimensions, in addition to length-based measures for overall complexity (see also Wolfe-Quintero et al., 1998).

Following Norris and Ortega (2009), recent investigations into grammatical complexity in L2 longitudinal corpora have employed several measures of clausal and phrasal complexity, in addition to length-based ones (Bulte & Housen, 2014 ; Kyle, 2016 ; Lu & Ai, 2015 ; Vyatkina, 2012 ; see also Cooper, 1976, for cross-sectional studies). These studies found evidence against the general developmental trajectories. In Bulte and Housen (2014), for example, L2 learners of English at an intermediate-advanced level showed increases in coordination (measured by the compound sentence ratio), phrasal elaboration (the mean length of finite clauses and of noun phrases) and overall length-based complexity (mean length of sentence, or MLS & MLT) at the end of four months of instruction. However, clausal complexity via subordination (the complex sentence ratio) remained the same. Similarly, Vyatkina (2012) found that overall sentence length, subordination and phrasal complexity measures increased as a course progressed, but that coordination measures decreased. In Kyle (2016), only MLT correlated positively with time in the analyses of two small-scale longitudinal corpora (Salsbury, 2000 ; Verspoor, Schmidt, & Xu, 2012).

One reason for the discrepancies in the research findings is the definition of a clause ; earlier studies only included finite clauses as clauses (see Wolfe-Quintero et al., 1998), while more recent research (Bulte & Housen, 2014 ; Vyatkina, 2012) included both finite and non-finite clauses. As Nakamura (submitted) pointed out, the number of dependent clauses increased in the latter approach, which might have led to the findings that were contrary to the general developmental trajectories.

Furthermore, as Nakamura (submitted) pointed out, the majority of these longitudinal investigations into L2 written production did not investigate direct relationships between overall-length measures and clausal and phrasal complexity indices. Without such investigations, we cannot know the dimensions (coordination, subordination or phrasal elaboration) to which learners add complexity when they lengthen texts. By contrast, several child language acquisition studies have investigated these relationships in spontaneous speech production (Blake, Quartaro, & Onorati, 1993 ; Hickey, 1991 ; Hunt, 1965 ; Scarborough, 1990 ; Scarborough, Rescorla, Tager-Flusberg, Fowler, & Sudhalter, 1991). What these studies showed were positive correlations between MLT in morphemes and clausal and phrasal complexity as measured via stage-wise developmental scales (Language Assessment, Remediation and Screening Procedure, Blake et al., 1993 ; Hickey, 1991), or via the total sum of various phrasal and clausal constructions (Index of Productive Syntax, Scarborough, 1990 ; Scarborough et al., 1991), particularly at an early phase of development (up to Mean Length of Utterances = 4.5 words). In SLA, only Kyle (2016) showed strong collinearities between overall length (MLT, MLS) and subordination measures (dependent clauses per clause), which indicated that learners composed longer sentences by adding more subordinate clauses. In a cross-sectional study of the grammatical complexity of essays written by adult L2 learners of English, Nakamura (submitted) found that there were significant positive correlations between length-based measures and the amount of subordination, as well as the length of a clause. Since there were also negative correlations between the amount of subordination and the length of a clause, the author argued that learners made longer sentences, or T-units, either by adding more dependent clauses or by lengthening a clause, but not both. This is in line with Hunt (1965).

Thus, one aim of this paper is to consider the development of overall length, clausal (coordination and subordination) and phrasal complexity, as well as the interrelations thereof.

Use of fine-grained measures

Another argument posited by Norris and Ortega (2009) was the possibility of the employment of developmentally sensitive, fine-grained or specific indices for clausal and phrasal complexity. In this regard, computational analysing tools for various textual features, such as grammatical complexity and cohesion, *Coh Metric* and the L2 Syntactic Complexity Analyzer (L2SCA) have gained in popularity in the literature on L2 writing (Bulte & Housen, 2014 ; Crossley & McNamara, 2012, 2014 ; Kyle, 2016 ; Lu & Ai, 2015 ; Mazgutova & Kormos, 2015). This is probably because these computational devices with fine-grained measures can capture developmental differences in the degrees of subcomponents (such as prepositional phrases or noun phrases) of a particular complexity dimension (such as phrasal elaboration). For example, MLC, an index of phrasal complexity suggested by Norris and Ortega (2009), cannot distinguish whether learners complexify a clause by adding longer nouns to a prepositional or subject phrase. Crossley and McNamara's (2014) study was one of the studies that used fine-grained measures. They employed the Coh Metric to analyse L2 essays written over a semester-long, intensive English programme (the same corpus as used in Bulte & Housen, 2014). The results showed that learners added more clauses, more modifiers in noun phrases or more words before the main verb and more negatives (*no*), but had less syntactic variety at the end of the semester. Mazgutova and Kormos (2015) found that learners in a higher proficiency group (IELTS mean score = 6.7) used more varied syntactic structure, while those in a lower proficiency group (mean score = 5.9) used fewer modifiers in noun phrases and simpler nominals, as well as more varied syntactic structures after a one-month intensive course for academic purposes. Neither group showed significant increases or decreases in overall complexity or clausal complexity by subordination measured via traditional, coarse-grained general indices such as MLT, or via dependent clauses per clause (DC per C). Nakamura (submitted) investigated whether the cloze test (see Thomas, 1994, for a review), one of the objective measures of proficiency could reliably capture the developmental differences in clausal and phrasal complexity. He showed that there were significant positive correlations between the scores on the cloze test and the amount of the dependents (modifiers) per nominal subject and objects, employing the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC) developed by Kyle (2016).

Based on his Cognition Hypothesis in Task-based Language Learning Literature, Robinson (Robinson, 2007; Robinson & Ellis, 2008) argued that *specific* indices, such as the use of psychological state terms (Robinson, 2007), target-like event conflations and the use of English aspectual systems (Robinson, Cadierno, & Shirai, 2009), in addition to general ones such as DC per C, should also be employed, since tasks differed in their cognitive complexity, which imposes different conceptual and cognitive demands (such as intentional versus visuo-spatial reasoning) on task implementation. In L2 writing domains, only a few studies have employed these "task-induced linguistic structures" (Kim & Tracy-Ventura, 2013) of *grammatical complexity* (see Johnson, 2017, for recent meta-analyses)[1].

These studies suggest that fine-grained complexity indices can capture subtle development with/without being promoted by tasks, which might be obscured by relying solely on general coarse-grained measures. Thus, another aim of this study is to track the development of phrasal complexity, particularly nominal complexity, by employing fine-grained measures provided by the TAASSC (Kyle, 2016), since phrasal complexity by nominalisation is a hallmark of advanced learners (Norris & Ortega, 2009) or a characteristic of an academic writing style (Crossley & McNamara, 2014).

Research Questions

RQ (1): How did overall, clausal and phrasal complexity develop over two semesters?

RQ (2): What relationships among overall, clausal and phrasal complexity held over two semesters?

RQ (3): Do fine-grained measures of nominal complexity show different developmental trajectories?

Method

Participants

Seventeen Japanese EFL freshmen (five male and 12 female) at an intermediate-advanced level university in Tokyo participated in this study[2]. Before participating in this study, they had learned English for at least six years, beginning in junior high school, as is typical in Japan. In the class, they learned English descriptive writing on various topics (such as writing a letter of recommendation for their favourite film) once a week over two semesters, based on the textbook (Kelly & Gargagliano, 2011). As part of the class, the following set of activities was provided: a) explanation + short exercises pertaining to the structure and contents of paragraphs, b) self-editing, and c) peer-editing. First and final drafts were assigned as homework. Thus, they were expected to write a paragraph at approximately three-week intervals and to produce eight paragraphs in total (see Table 1 for short descriptions of each assignment). Only first drafts were analysed in this paper, and all direct quotations were removed from the texts.

Grammatical complexity indices

The following three types of measures were employed in the analyses: a) overall complexity by length, b) clausal (coordination and subordination) complexity and c) phrasal complexity, using Lu's (2010) L2SCA (Table 2).

In addition, as fine-grained measures of nominal complexity, the following seven measures provided in Kyle's (2016) TAASSC (Table 3) were used to investigate RQ (3)[3 4].

'Dependents' here refer to any dependent elements (such as adjectives) of nominal; thus, these seven fine-grained indices measure the number of modifying elements that learners added to each nominal type.

Table 1 Topics and Contents of the Writing Tasks

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
Topic	An important place	An ideal partner	A favourite photo	Favourite art	Party Time	Invitation letter	Film review	My favourites/ my hometown
Contents	Describe an important place in your life	Describe an ideal partner	Describe a favourite photo	Describe a favourite artwork	Write an invitation letter to your friend	Write a thank you note to your friend	Choose one favourite film and write a review	Describe either a) your one of your favourites (other than those written about in Tasks 3 and 4, or b) your hometown

Note. Tasks 4 and 8 were written during the summer and winter breaks, respectively.

Table 2 Complexity Indices Taken from Lu (2010)

Types of Complexity Indices			
Overall Complexity		Mean Length of T-unit (MLT)	The number of words divided by that of T-units
Clausal Complexity	Coordination	T-units per Sentence (T per S)	The number of T-units divided by that of sentences
	Subordination	a) Clauses per T-unit (C per T)	The number of clauses divided by that of T-units
		b) Dependent Clauses per Clause (DC per C)	The number of dependent clauses divided by that of clauses
		c) Dependent clauses per T-unit (DC per T)	The number of dependent clauses divided by that of T-units
Phrasal complexity		Mean Length of Clauses (MLC)	The number of words divided by that of clauses

Table 3 Fine-grained Indices of Nominal Complexity Taken from Kyle (2016)

Types of indices	
Dependents per Nominal (D per N)	The number of dependent elements divided by that of nominals
Dependents per Subject Nominal (D per SN)	The number of dependent elements divided by that of nominal subjects
Dependents per Passive Nominal Subject (D per PNS)	The number of dependent elements divided by that of passive nominal subjects
Dependents per Direct Object (D per DO)	The number of dependents elements divided by that of direct objects
Dependents per Object of the Preposition (D per OP)	The number of dependents elements divided by that of object nominals of prepositions
Dependents per Indirect Object (D per IO)	The number of dependent elements divided by that of indirect objects
Dependents per Nominal Complement (D per NC)	The number of dependent elements divided by that of nominal complements

Results

Means, SDs, and 95% CIs are available from the author's *ResearchGate* (https://www.researchgate.net/publication/330982776_Online_Supplementary_Materials_Nakamura_2019). Since some of the data did not show normal distributions, the data were analysed using the nonparametric Friedman ANOVAs for RQ (1) and (3), and multiple regression analyses with log-transformation of MLT and Kendall's τ between C per T and MLC for RQ (2) (see discussions pertaining to the inclusion of Kendall's τ).

RQ (1): How did overall, clausal, and phrasal complexity develop over two semesters?

As Figure 1 shows, there were two discernible fluctuation patterns: a) except for MLC, the measure of phrasal complexity, all measures seemed to show M-shaped developmental curves (repeated rise-then-down patterns) and b) MLC apparently showed a reversed M- (or W) shaped curve.

However, a series of Friedman ANOVAs in which time served as an independent variable and subsequent paired comparisons revealed that significant differences appeared at different time points (see Table 4). Firstly, the overall length (MLT) showed a significant, early developmental peak at time 4 (time 1＜time 4), and followed by a decrease at time 5 (time 5＜time 4).

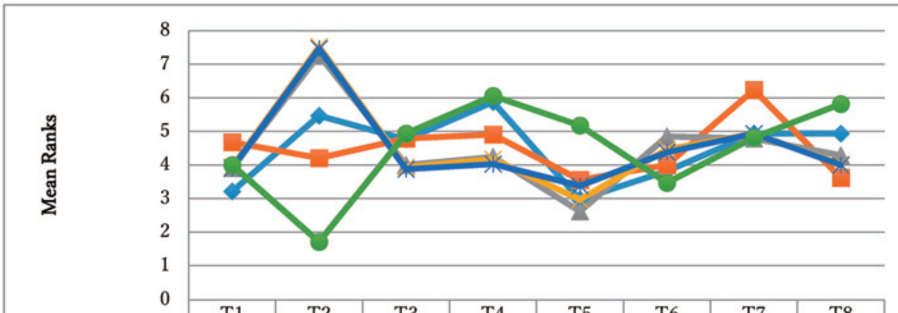
Coordination, as indexed by T per S, demonstrated the slowest developmental peak at time 7 (time 5＜time 7).

Subordination, measured by C per T, DC per T, and DC per C, showed peaks at time 2 (time 1＜time 2), the earliest developmental peak amongst the clausal and phrasal measures, and then fell at time 3 (time 3＜time 2).

Finally, phrasal complexity measured by MLC showed the second earliest developmental peak at time 3 (time 2＜time 3), and then remained unchanged. In other words, the following developmental trends were observed: from subordination through phrasal elaboration to coordination, which contradicts the general developmental trajectories observed in the literature (Norris & Ortega, 2009; Wolfe-Quintero et al., 1998).

RQ (2): What relationships among overall, clausal and phrasal complexity held over two semesters?

In order to address RQ (2), multiple regression analyses were conducted with log-transformation of the MLT scores. C per T was chosen as a representative measure of subordination in order to avoid multicollinearity, since three subordination indices



	T1	T2	T3	T4	T5	T6	T7	T8
MLT	3.21	5.47	4.76	5.88	2.97	3.82	4.94	4.94
T per S	4.68	4.21	4.79	4.91	3.56	4	6.24	3.62
C per T	3.91	7.29	4	4.24	2.62	4.85	4.79	4.29
DC per T	3.94	7.53	3.94	4.18	2.97	4.47	4.94	4.03
DC per C	3.91	7.47	3.88	4.03	3.38	4.38	4.94	4
MLC	4	1.71	4.94	6.06	5.18	3.47	4.82	5.82

Figure 1　Mean Rank Orders of Six Grammatical Complexity Indices

Table 4 Results of Friedman ANOVAs and Paired Comparisons

	MLT	T per S	C per T	DC per T	DC per C	MLC
Friedman ANOVAs	$X^2=22.07$ (p=.002)	$X^2=16.09$ (p=.024)	$X^2=34.96$ (p=.000)	$X^2=36.08$ (p=.000)	$X^2=32.69$ (p=.000)	$X^2=39.82$ (p=.000)
Paired comparisons	T1<T4 (p=.04) T4>T5 (p=.015)	T5<T7 (p=.040)	T1<T2 (p=.002) T2>T3 (p=.002) T2>T4 (p=.008) T2>T5 (p=.00) T2>T8 (p=.010)	T1<T2 (p=.001) T2>T3 (p=.001) T2>T4 (p=.002) T2>T5 (p=.00) T2>T6 (p=.008) T2>T8 (p=.001)	T1<T2 (p=.001) T2>T3 (p=.001) T2>T4 (p=.001) T2>T5 (p=.00) T2>T6 (p=.007) T2>T8 (p=.001)	T2<T3 (p=.003) T2<T4 (p=.00) T2<T5 (p=.001) T2<T7 (p=.006) T2<T8 (p=.00)

Note. All *p* values in paired comparison columns are adjusted *p* values.

always showed higher positive intercorrelations. T per S, C per T, and MLC were taken as predictor variables and MLT as a dependent variable. None of the predictors showed multicollinearity (VIF=< 10 and Tolerance statistics>.02, Field, 2013). Table 5 provides the results of the multiple regression analyses.

Firstly, contributions by T per S, a coordination index, to the overall length were negligible as the β scores were quite small and nonsignificant (and the βs always crossed zero). Secondly, C per T and MLC showed the largest contributions interchangeably (compare βs): C per T showed the largest contributions to the overall length initially (up to Time 3) and MLC then took that position at time 4. At time 5, the β scores of the C per T became the largest again and, finally, MLC showed the largest β at time 8.

RQ (3): Do fine-grained measures of nominal complexity show different developmental trajectories?

As the descriptive statistics provided in Figure 2 showed, except for D per IO, a measure of the average number of dependents per indirect object, all measures seemed to show a W-shaped developmental pattern with peaks at time 4 and times 7 or 8 (Means, SDs, 95% *CI* are available from the author's *ResearchGate* (https://www.researchgate.net/publication/330982603_Online_Supplementary_Materials2_Nakamura2019).

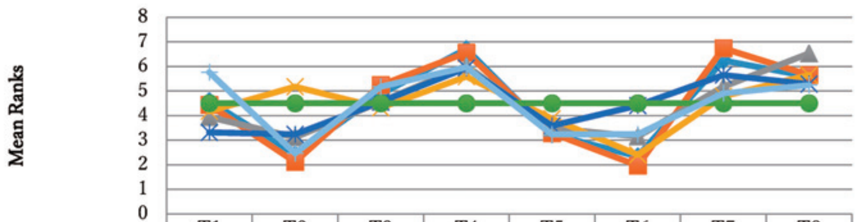
Furthermore, the Freidman ANOVAs revealed that time had significant effects on the measures except for D per IO, as Table 6 shows. However, Table 6 also indicates that each measure had different developmental trajectories.

Firstly, D per N, D per NS, and D per PNS showed an N-shaped developmental pattern; learners attached more modifying elements such as adjectives to nominals in general, or to subject nominals in particular at time 3 or time 4, followed by fewer at times 5 or 6, and more again at times 7 or 8. Secondly, learners added more modifying elements to direct objects (D per DO) initially, followed by fewer at time 6, and more again at time 8 (thus producing a V-shaped pattern). Thirdly, learners added more modifications to object nominals of prepositions at time 4 (D per OP). Fourthly, learners initially reduced the number of modifications on nominal complements at time 2, then added more at time 3, and reduced them again (a mirrored N-pattern). Finally, learners did not add or reduce the number of modifying elements on to indirect objects, as revealed by the null effects of time on this measure ($X^2=.000$).

Table 5　Results of Multiple Regression Analyses and Kendall's τ between C per T and MLC

Dependent Variables	Predictor variables	B (95% CI)	SE B	β	p	Kendall's τ
MLT1[ab]	Constant	.811 (.655, .968)	.074		.000	−.379
	C per T1	.139 (.026, .252)	.053	.561	.019	p=.038
MLT2[c]	Constant	.234 (.160, .309)	.035		.000	−.127
	MLC2	.067 (.059, .075)	.004	.523	.000	ns.
	C per T2	.214 (.199, .230)	.007	.884	.000	
	T per S2	.006 (−.049, .061)	.025	.007	ns.	
MLT3[d]	Constant	.254 (.122, .387)	.061		.001	.075
	MLC3	.048 (.830, .997)	.003	.542	.000	ns.
	C per T3	.265 (.236, .294)	.013	.776	.000	
	T per S3	.021 (−.086, .128)	.050	.017	ns.	
MLT4[e]	Constant	.209 (.156, .262)	.025		.000	−.081
	MLC4	.046 (.043, .049)	.002	.809	.000	ns.
	C per T4	.293 (.270, .317)	.011	.707	.000	
	T per S4	.048 (.003, .092)	.021	.062	.038	
MLT5[f]	Constant	.122 (.004, .241)	.055		.044	−.284
	MLC5	.052 (.047, .057)	.002	.839	.000	ns.
	C per T5	.339 (.297, .382)	.020	.643	.000	
	T per S5	.024 (−.050, .098)	.034	.026	ns.	
MLT6[g]	Constant	.196 (.126, .267)	.033		.000	−.262
	MLC6	.060 (.054, .066)	.003	.761	.000	ns.
	C per T6	.260 (.239, .280)	.009	.955	.000	
	T per S6	.006 (−.038, .050)	.020	.010	ns.	
MLT7[hi]	Constant	.275 (.231, .320)	.021		.000	−.124
	MLC7	.049 (.044, .053)	.002	.616	.000	ns.
	C per T7	.269 (.248, .290)	.010	.721	.000	
MLT8[k]	Constant	.561 (.073, 1.049)	.226		.027	.015
	MLC8	.032 (.004, .061)	.013	.552	.030	ns.
	C per T8	.089 (−.042, .220)	.061	.296	ns.	
	T per S8	.082 (−.377, .542)	.213	.087	ns.	

Note. a: adjusted R^2=.269, p=.019. b: Initial multiple regression analysis included MLC2 but showed β beyond 1 for both C per T2 and MLC2. Removal of T per S from the model did not solve the problem. Therefore, the anomaly was due to the inclusion of MLC2. A simple regression with MLC confirmed this: The model did not fit the data (adjusted R^2=.007, p=.308). c: adjusted R^2=.987, p=.000. d: adjusted R^2=.976, p=.000. e: adjusted R^2=.991, p=.000. f: adjusted R^2=.979, p=.000, ΔR=.388 (Step 2), ps=.000. g: R^2 (Step 1)=.496. p=.002, ΔR (Step 2)=.504, p=.000. h: adjusted R^2=.989, p=.000. i: Initial multiple regression showed negative values on the standardized coefficient of T per S. Thus T per S was excluded from the model. k: adjusted R^2=.366, p=.030.



	T1	T2	T3	T4	T5	T6	T7	T8
D per N	4.59	2.41	4.88	6.71	3.24	2.35	6.24	5.59
D per NS	4.44	2.12	5.24	6.56	3.29	1.97	6.74	5.65
D per PNS	3.94	3.12	4.56	6.03	3.5	3.15	5.18	6.53
D per DO	4.18	5.18	4.32	5.59	3.88	2.41	4.79	5.65
D per OP	3.32	3.24	4.56	5.94	3.59	4.41	5.65	5.29
D per IO	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5
D per NC	5.76	2.47	5.18	5.97	3.24	3.24	4.91	5.24

Figure 2　Mean Rank Orders of Seven Nominal Complexity Indices

Table 6　Results of Friedman ANOVAs and Paired Comparisons of Nominal Complexity

	D per N	D per NS	D per PNS	D per DO	D per OP	D per IO	D per NC
Friedman ANOVAs	X^2=56.059 p=.000	X^2=70.207 p=.000	X^2=44.481 p=.000	X^2=22.588 p=.002	X^2=22.397 p=.002	X^2=.000 ns.	X^2=36.065 p=.000
Paired comparisons	T2<T4 (p=.000) T2<T7 (p=.000) T2<T8 (p=.004) T4>T5 (p=.001) T4>T6 (p=.000) T5<T7 (p=.010) T6<T7 (p=.000) T6<T8 (p=.003)	T2<T3 (p=.006) T2<T4 (p=.000) T2<T7 (p=.000) T2<T8 (p=.001) T3>T6 (p=.003) T4>T5 (p=.003) T4>T6 (p=.000) T5<T7 (p=.001) T6<T7 (p=.000) T6<T8 (p=.000)	T2<T4 (p=.015) T2<T8 (p=.001) T4>T6 (p=.017) T5<T8 (p=.009) T6<T8 (p=.002)	T2>T6 (p=.028) T4>T6 (p=.004) T6<T8 (p=.003)	T2<T4 (p.=.036)	ns.	T1>T2 (p=.002) T2<T3 (p=.036) T2<T4 (p=.001) T2<T8 (p=.028) T4>T5 (p=.032) T4>T6 (p=.032)

Discussions

Development trajectories of grammatical complexity by length, clauses and phrases

The present study found that the earliest developmental peak was observed in subordination (at time 2), followed by phrasal elaboration (at time 3) and general overall length (at time 4). Coordination showed the slowest developmental peak (at time 7); thus, the trend observed here is from subordination through phrasal elaboration (and overall length) to coordination (see Table 4 and Figure 1). This is contradictory to both the general trends from coordination through subordination to phrasal elaboration found in the literature (Norris & Ortega, 2009 ; Wolfe-Quintero et al., 1998), and to recent longitudinal research on L2 writing (Bulte & Housen, 2014).

One of the reasons for such discrepancies is methodological ; while some studies (Bulte & Housen, 2014 ; Crossley & McNamara, 2014) compared only the initial with the final time points, the present study analysed all the data points over one year. If we compare only the first with the last writing, none of the measures shows development, as indicated in the lack of significant differences between time 1 and time 8 (see Table 4). Therefore, the former approach might hide the potential developmental differences in coordination, subordination and phrasal elaboration at the intermediate time points as observed in the present study. In other words, comparisons of initial with final writing may provide quasi (null) development of grammatical complexity. Moreover, as noted above, recent research (Bulte & Housen, 2014 ; Kyle, 2016 ; Vyatkina, 2012) included both non-finite and finite clauses as examples of clauses, while the present study and representative research synthesis (Wolfe-Quintero et al., 1998) reserve a clause only for a finite one. In particular, the former approach may have inflated the number of subordinations.

The question then is why the results of the present study, particularly the slowest developmental peak of coordination, contradicted the general developmental trends, which indicated that coordination should develop initially. The results for RQ (2) might shed light on this.

Coordination, subordination, phrases and their relationships to overall length

As Table 5 shows, at the initial developmental phase, only subordination (C per T1) contributed to the overall length. After fluctuations in the magnitude of contributions between subordination and phrasal elaboration (compare the βs of C per T 2-7 with those of MLC2-7) in the intermediate phases, only phrasal elaboration (MLC8) contributed to the overall length of the text in the final developmental phrase. What these results suggest is that learners initially constructed their texts by adding more subordinate clauses, and that they finally did so by adding more modifying elements to a clause. Coordination had no significant contribution to the overall length at all.

If we assume that the general developmental trends observed in the literature were more concerned with the *overall length* than they were with the *time* of the course of learning, the picture becomes clearer. Learners in the present study were taking the course at the advanced level of proficiency. At the initial intermediate proficiency level, subordination is a responsible index, and thus showed the largest βs. At the final advanced proficiency level, phrasal elaboration is the factor responsible for development, and thus showed the largest βs. Fluctuations in the magnitudes of the βs of adjacent predictors (subordination and phrasal elaboration) may represent the transition to the next developmental stage. To confirm this interpretation, additional nonparametric Kendall's τ correlational analyses between C per T and MLC were conducted. Table 5 shows that, as in Nakamura (submitted) and Vyatkina (2012), the relationships between subordination and phrasal elaboration were competitive, as the correlations between them are always negative (although most of them are nonsignificant). This means that learners lengthened their texts either by adding more subordinate clauses or by making a clause longer, but not both. Although coordination showed the slowest improvements, this had no relationship with the overall length, as suggested by the smallest and often non-significant βs in the regression models. It may be that such improvements, shown in the Friedman ANOVAs with time as an independent variable, have no direct relationship with the general developmental trends *per se*, since the learners in the present study had already passed the early phase of development in which complexity by coordination should play a role (Norris & Ortega, 2009).

One possible modulating factor is the effect of genre. As shown in Table 1, the present paper employed several different topics : some were purely descriptive (such as Task 5), while others were descriptive+argumentative (such as Task 2). While this was inevitable because the research was tied closely to Kelly and Gargagliano (2011), it has been pointed out that different genres (Mazgutova & Kormos, 2015) or "task types" in general (Skehan, 2009) have different effects on L2 production. Thus, the consistent use of particular genres over a semester and investigations into their respective effects on L2 written production are issues for further research (see Nakamura, 2018, for a preliminary investigation).

Differences in the nominal modifications of each subtype

The present study found that learners added modifications to nominals at different time points, except for nominals in the indirect object slot (see Figure 2 and Table 6). However, some commonalities were also found ; nominals in general (D per N) and subject nominals in particular (D per NS & D per PNS), as well as one in the direct object slot, showed initial increases through intermediate declines at time 6 to a final rise in the number of modifying elements. No studies have investigated the different development of each nominal type, although general trends of increases in modifications of nominals found here support the findings of Bulte and Housen (2014) and Mazgutova and Kormos (2015) ; however, as adequate reasons for such developmental differences cannot be provided here, the present study is exploratory in nature. Tracking the developmental trajectories of each nominal type, particularly its type and token frequencies in language input as well as output, as Usage-based models of language acquisition (Tomasello, 2003 ; Robinson & Ellis, 2008) have done, will pave the way for clarifying the nature of the micro-development (Grannot & Parziale, 2009) of L2 grammatical complexity. If a particular "type" appears more frequently in the input than do other types and the output reflects the frequency distributions in the input, then frequency effects might be one possible reason for developmental differences. Accumulations of research findings on these subdimensions of a particular dimension of grammatical complexity, particularly nominal complexity, are also necessary since nominalisation is the hallmark of advanced L2 capacity (Norris & Ortega, 2009) or evidence of the acquisition of an academic writing style (Crossley & McNamara, 2014).

Conclusion

The present study, by employing multiple indices of overall length, clausal and phrasal complexity, found that L2 learners developed these dimensions at different speeds. If we track their development across all data points, and do not simply make comparisons of the initial and final learning in their writing, the results will indeed follow the general developmental trajectories from coordination through subordination to phrasal elaboration based on the assumption that the general developmental trajectories are concerned with the overall length of the text that learners produce instead of the duration of the course of learning. In terms of nominal complexity analysed via fine-grained measures, each nominal subtype showed different developmental trajectories with commonalities in that there were initial and final increases in the number of modifying elements with reductions in the middle phase for nominals in general and for one in subject and direct object slots in particular. More research on the subdimensions of a particular dimension is clearly necessary to clarify the macro development of grammatical complexity.

One limitation of the present study is the lack of independent proficiency measures. The present study extrapolated learners' English proficiency from their university level. Although this is one typical way to measure learners' proficiency in the literature on SLA (see Thomas, 1994), some independent measures such as the cloze test are clearly necessary to gauge their proficiency[5].

Notes

1 I could not develop and employ task-induced indices

of grammatical complexity since the learners in this study wrote essays on a different topic in each time phase; thus, these indices could not capture their development over a year. This task sequencing (Baralt, Gilabert, & Robinson, 2014), or task repetition (Bygates, 2001), deserves attention in future studies.

2 One reviewer pointed out the possibility of gender differences. Since the number of the participants was rather small and thus statistically reliable analyses could not be given, this would be one issue for future research.

3 TAASSC contains 31 indices of clausal complexity and 132 measures of phrasal complexity. I only included these seven measures of phrasal complexity since

a) the TAASSC identifies both finite and non-finite clauses as clauses, while only the former type is considered to be a clause in the L2SLA; thus, the results are not comparable, and

b) the aim of the present paper is to investigate the development of phrasal complexity of basic grammatical categories (subject, direct, indirect object, object of prepositions and compliment) as in Nakamura (submitted).

To investigate further finer-distinctions of nominal complexity (such as adjectival modifiers, prepositions or possessives per nominal subject), much longer essays than the paragraphs targeted in this study seem to be required.

4 Of these, there were only a few examples of D per IO and, since this measure had no significant differences, data on D per IO were not presented in the remainder of the paper.

5 Originally, the participants in the present study completed a cloze test, as in Nakamura (submitted). However, the data were lost because of the author's mistake.

References

Arthur, B. (1979). Short-term changes in EFL composition skills. In G. Yorio, K. Perkins, & J. Schacter (Eds.), *On TESOL '79 : The Learner in Focus* (pp.330-342). Washington, DC : TESOL.

Baralt, M., Gilabert, R., & Robinson, P. (Eds.). (2014). *Task sequencing and Instructed Second Language Learning*. London : Bloomsbury.

Bygates, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygates, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks : Second Language Learning and Testing* (pp.23-48). Harlow : Longman.

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language, 20*, 139-152.

Bulte, B. & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42-65.

Casanave, C.P. (1994). Language development in students' journals. *Journal of Second Language Writing, 3* (3), 179-201.

Cooper, T.C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research, 69* (5), 176-183.

Crossley, S.A. & McNamara, D.S. (2012). Predicting second language writing proficiency : The roles of cohesion and linguistic sophistication. *Journal of Research in Reading, 35* (2), 115-135.

Crossley, S.A. & McNamara, D.S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26*, 66-79.

Filed, A. (2013). *Discovering Statistics using IBM SPSS Statistics*. Fourth edition. London : Sage.

Grannot, N. & Parziale, J. (Eds.). (2009). *Microdevelopment : Transition Processes in Development and Learning*. Cambridge : Cambridge University Press.

Hickey, T. (1991). Mean length of utterance and the acquisition of Irish. *Journal of Child Language, 18*, 553-569.

Hogue, A. (2008). *First Steps in Academic Writing*. Second Edition. New York : Pearson.

Hunt, K.W. (1965). Grammatical structures written at three grade levels. National Council of Teachers of English. *NCTE Research Report No. 3*.

Johnson, M.D. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency : A research synthesis and meta-analysis. *Journal of Second Language Writing, 37*, 13-38.

Kelly, C. & Gargagliano, A. (2011). *Writing from within. Level 1*. Second edition. New York : Cambridge University Press.

Kern, R.G. & Schultz, J.M. (1992). The effects of composition instruction on intermediate level French students' writing performance : Some preliminary findings. *The Modern Language Journal, 76*, 1-13.

Kim, Y. & Tracy-Ventuna, N. (2013). The role of task repetition in L2 performance development : What needs to be repeated during task-based interaction? *System, 41*, 829-840.

Kyle, K. (2016). *Measuring syntactic development in L2 writing : Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*.

Unpublished doctoral dissertation. George Town University.

Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H.W. Seliger & M.H. Long (Eds.), *Classroom-Oriented Research in Second Language Acquisition* (pp.287-304). Rowley, MA: Newbury House.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27* (4), 590-619.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*, 474-496.

Lu, X. & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing, 29*, 16-27.

Mazgutova, D. & Kormos, J. (2015). Syntactic and lexical development in an intensive English for academic purposes programme. *Journal of Second Language Writing, 29*, 3-15.

Nakamura, D. (2018). *Effects of task repetition on the development of grammatical complexity in L2 summary writing*. Paper presented at TBLT in Asia, the fourth biennial conference presented by the JALT TBL SIG, June 23 and 24, Ryukoku University, Kyoto.

Nakamura, D. (submitted). *The cloze test and overall, clausal, and subclausal indices of grammatical complexification in second language essays: A cross-sectional study of the developmental trajectories*.

Norris, J.M. & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30* (4), 555-578.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24* (4), 492-518.

Ortega, L. & Byrnes, H. (2008a). *The Longitudinal Study of Advanced L2 Capacities*. New York: Routledge.

Ortega, L. & Byrnes, H. (2008b). The longitudinal study of advanced L2 capacities: An introduction. In L. Ortega & H. Byrnes (Eds.), *The Longitudinal Study of Advanced L2 Capacities* (pp.3-20). New York: Routledge.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching, 45*, 191-213.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and motion: Measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics, 38*, 533-554.

Robinson, P. & Ellis, N.C. (2008). Conclusion: Cognitive linguistics, second language acquisition and L2 instruction-Issues for research. In P. Robinson & N.C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp.489-545). New York: Routledge.

Salsbury, T. (2000). *The acquisition of grammaticalization of unreal conditionals and modality in L2 English: A longitudinal perspective*. Indiana University.

Scarborough, H.S. (1990). Index of productive syntax. *Applied Psycholinguistics, 11*, 1-22.

Scarborough, H.S., Rescorla, L., Tager-Flusberg, H., Flower, A.E., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics, 12*, 23-45.

Skehan, P. (2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics, 30* (4), 510-532.

Thomas, M. (1994). Assessment of L2 Proficiency in Second Language Acquisition Research. *Language Learning, 44* (2), 307-336.

Tomasello, M. (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Verspoor, M., Schmid, M.S., & Xu, X. (2012). A dynamic usage-based perspective on L2 writing. *Journal of Second Language Writing, 21* (3), 239-263.

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *Modern Language Journal, 96* (4), 576-598.

Wolfe-Quintero, K., Inagaki, S., & Kim, H-Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, & Complexity*. Honolulu, HI: University of Hawai'i Press.

(2018年 9 月 6 日受付
2018年12月12日受理)