

2024 年度『日本語教育』論文賞受賞論文—概要—
Research Paper Summary for the 2024 Best Paper Award
of Association for Japanese Language Education

**Quantitative Study on the Complexity in Sentence Structures of
Japanese Language Learners:
An Analysis of I-JAS Story Writings**

LI Wenping and LIU Haitao

Complexity of sentence structures is a key research focus in language education. In Japanese language education, the complexity of learners' sentence structures has been investigated from two dimensions: "horizontal expansion" and "vertical embedding." In recent years, "T-units per sentence" (hereafter T/S) and "clauses per T-unit" (hereafter C/T) have been employed as indicators of horizontal expansion and vertical embedding, respectively. However, these measures fail to distinguish horizontal expansion across proficiency levels, nor can they capture the interaction between horizontal expansion and vertical embedding.

To more accurately assess the complexity of Japanese language learners' sentence structures, this study quantitatively analyzed features related to the complexity of sentence structures in terms of both horizontal expansion and vertical embedding as learners' proficiency improves. The analysis examined 300 story writings from the International Corpus of Japanese as a Second Language (I-JAS), using the metrics of mean dependency distance (MDD) and sentence structure characteristic value (SSCV) to quantify how complexity changes.

Effectiveness of MDD and SSCV

The effectiveness of MDD and SSCV is examined by comparing them with traditional measures (T/S and C/T). With respect to horizontal expansion, MDD measures complexity more precisely than T/S. While T/S can capture differences in horizontal expansion between sentences with different numbers of T-units, it cannot distinguish differences between sentences that contain the same number of T-units. In contrast, MDD can capture such distinctions. In the examples below, the underlined parts represent subordinate clauses and the parts in square brackets represent T-units. Example (2) is more complex than Example (1) in terms of horizontal expansion, yet comparable to Example (3). Whereas T/S fails to reflect these differences and similarities, MDD captures both.

(1) [Doa o akeru koto ga dekimasen deshita.]

Ken could not open the door.

【TTR08-SW2 1T-unit subordinate clause:1 T/S:1 MDD:1.0】

(2) [*Ken wa kagi o wasureta node, uchi ni hairemasen deshita.*]

Ken forgot his key, so he could not enter the house.

【SES02-SW2 1T-unit subordinate clause:1 T/S:1 MDD:1.5】

(3) [*Ken wa Mari to yobimasu ga,*] [*Mari wa nete imasu.*]

Ken calls out to Mari, but Mari is asleep.

【SES22-SW2 2T-unit subordinate clause:0 T/S:2 MDD:1.5】

Similarly, for vertical embedding, SSCV offers a finer measure than C/T. Although C/T can capture differences in vertical embedding between sentences with different numbers of subordinate clauses, it cannot distinguish differences between sentences that contain the same number of subordinate clauses. In Examples (4), (5), and (6), the number of clauses is calculated as the sum of main clauses and subordinate clauses. In other words, a sentence without subordinate clauses contains only one clause. Example (5) has a structurally more complex form than Example (4), and Example (6) is comparable to (5). While C/T does not reflect such differences or similarities, SSCV successfully captures both.

(4) [*Sono toki wa heya ni inu ga imashita.*]

At that time, there was a dog in the room.

【RRS10-SW1 1T-unit subordinate clause:0 C/T:1 SSCV:1.0】

(5) [*Inu wa basuketto no naka no tabemono o tabete shimaimashita.*]

The dog ate the food that was in the basket.

【GAT03-SW1 1T-unit subordinate clause:0 C/T:1 SSCV:1.4】

(6) [*Sore o shiranai futari wa pikunikku o tanoshinde mashita.*]

The two people, unaware of that, enjoyed their picnic.

【KKD16-SW1 1T-unit subordinate clause:1 C/T:2 SSCV:1.4】

Syntactic complexity across proficiency levels

The study further analyzes changes in complexity across proficiency levels. First, vertical embedding shows greater variation than horizontal expansion, as indicated by the larger effect size of SSCV compared to MDD. This suggests that vertical embedding changes more prominently with proficiency. Second, MDD increases significantly between the N5-N4 and N3-N2 stages (N5-N4: $p = 0.001$; N3-N2: $p = 0.000$), but the difference between N2 and N1 is not significant ($p = 0.111$). This indicates that horizontal expansion slows at higher proficiency. Third, SSCV increases significantly across all stages (N5-N4: $p = 0.000$; N3-N2: $p = 0.000$; N2-N1: $p = 0.016$), with medium effect size at N5-N4 and N3-N2, and small effect size at N2-N1. This suggests that vertical embedding also shows a diminished increase at higher stages.

Applications of MDD and SSCV

Compared with T/S and C/T, MDD and SSCV provide more accurate measures of both

horizontal expansion and vertical embedding. These indices have at least three applications in Japanese language education: (1) assessing learner proficiency, evaluating learning outcomes, and informing syllabus design; (2) providing new quantitative metrics to identify structural sources of reduced comprehensibility in learner writing; and (3) serving as effective indicators of sentence structure to address the current lack of structural metrics in readability tools, thereby supporting the development of automatic readability assessment systems.

(LI: Shanghai University of Finance and Economics, LIU: Zhejiang University)

(published in the April 2024 edition of the Association's *Journal of Japanese Language Education*)