International Journal of Education, Learning and Development
Vol. 10, No.9, pp.55-65, 2022
Print ISSN: 2054-6297(Print)
Online ISSN: 2054-6300 (Online)

Turkish Japanese learners' L3 Japanese oral proficiency

Wenchao Li
Zhejiang University

Citation: WenchaoW Li (2022) Turkish Japanese learners' L3 Japanese oral proficiency, *International Journal of Education, Learning and Development*, Vol. 10, No.9, pp.55-65

ABSTRACT: *This study applies mathematic linguistics to explore Turkish Japanese learners' oral proficiency. Data were drawn from a self-built annotated corpus of 300 hours of recordings of storytelling (individual) and interaction (group). The oral proficiency for lexical sophistication is measured utilising moving-average morphological richness (MAMR) and mean size of paradigm. The syntactic complexity is measured by mean dependency distance (MDD). Self-written computer programme scripts are used to compute the MAMR, moving-average mean size of paradigm (MAMSP), and MDD. The findings indicate that in both storytelling and interaction, Turkish Japanese learners present a close picture regarding the degree of freedom of MAMR, MAMSP, and MDD to the oral data of native Japanese. This study, therefore, contends that L1 mother tongue morpho-syntactically affects later language acquisition. Moreover, the dependency distance (DD) that attributes the most tokens in interaction is nearly half that of the DD in storytelling, confirming the Principle of Least Effort (Zipf 1949) that human action tends to lighten the processing load. Furthermore, the MDD-frequency relationship in the Turkish oral data demonstrates a good fitting result for three models, that is, Harris, Dacey, and Kelly.*

KEYWORDS: Japanese language acquisition, oral proficiency, lexical sophistication, syntactic complexity

INTRODUCTION

Also known as syntactic maturity, syntactic complexity refers to the use of a range of forms with degrees of sophistication (Ortega 2003). It has been regarded as a valid and reliable metric in examining language acquisition quality (Iwashita 2006) and has been extensively employed in cross-linguistic studies, for example, Hunt (1970), Witte and Sodowsky (1978), Falhive and Snow (1980), Wolfe-Quintero et al. (1998). Existing studies seem to mostly focus on second language acquisition, particularly English, with third and second languages other than English relatively underexplored.

Previously adopted measures include T-unit (Harrington 1986) and type-token (Richards 1987). Recently, it has been proved that mean dependency distance (MDD), a concept under dependency grammar, is able to reflect the acquisition proficiency, for example, Jiang and Liu (2015), Jiang, Bi, and Liu (2019), Komori et al. (2019), Li and Yan (2021). To further confirm the link between oral proficiency and morphosyntactic complexity, this study delves into Turkish Japanese learners' oral proficiency from a mathematical linguistics viewpoint. Given that Turkish and Japanese have undeniable similarities regarding morphology and syntax, the study outcomes may also shed light on the association between language distance and acquisition.

The Turkish language is a member of the Altaic language family and is highly agglutinative. The case system is marked by suffixes, as shown in (1).

(1) oku-ma-y-abil-ir-im
read（stem）-[Neg]-(inserted consonant)-[possibility]- [present tense]-[1st sg]
'I may not read.'

Essentially, case markers obey vowel harmony as shown in (2) and vowel-consonant harmony as shown in (3).

(2) **Turkish dative case affixes (vowel harmony)**

(i) /-a/: attach to the lexemes that end with vowels /a/, /ı/, /o/, /u/, e.g. *okula* (go to).

(ii) /-e/: attached to the lexemes that end with vowels /e/, /i/, /ö/, /ü/, e.g. *işe* (drive to).

(3) **Turkish accusative case affixes (vowel-consonant harmony)**

(i). /-yı/, /-yi/, /-yu/, /-yü/: attach to lexemes that end in vowels, e.g. *gazeteyi* (being reading).

(ii). /-ı/, /-i/, /-u/, /-ü/: attach to lexemes that end in consonants, e.g. *sandviçi* (want).

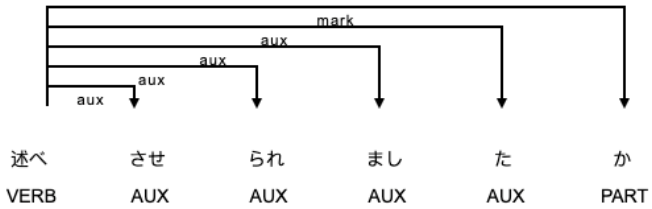
Although Japanese affiliation remains under discussion, there is no doubt that morphologically Japanese is agglutinative with one or more suffixes being added to a verb/adjective stem to give rise to complex predicates, for example (4). (5) is a syntactic structure for the illustration of (4).

(4) **agglutinative**

述べ-させ-られ-まし-た-か

nobe-sase-rare-mashi-ta-ka.
speak(stem)-causative-passive voice-honorification-tense.past-question marker

(5)



In Old Japanese (AD 700–800; a dead language spoken in Asuka and Nara periods), there are eight vowels, that is, /a/, /e_1/, /e_2/, /i_1/, /i_2/, /o_1/, /o_2/, and /u/. This leads to the deduction that harmony existed thousand years ago. The similarities between Turkish and Japanese are further tied to word order; the basic word order for the two languages is subject-object-verb (SOV).

Despite the undeniable similarities in morphology and syntax, the two differ in case marking systems. The Turkish case is marked by a suffix, whilst the Japanese case is rendered by a particle. At this stage, a question arises as to how Turkish Japanese learners would perform when speaking Japanese; would typological distance play an essential role in L3 Japanese acquisition? This study employs the moving-average morphological richness (MAMR) and the moving average mean size of paradigm (MAMSP) to examine the lexical sophistication. The mean dependency distance (MDD) is employed to measure the syntactic complexity utilising data of 300 hours of recordings of storytelling (individual) and conversation (group). We aim to derive an understanding of Turkish Japanese learners' oral proficiency. We also aim to determine whether language closeness between the mother tongue and later acquired language may affect the learning quality of the latter.

The rest of this paper is organised as follows. Section 2 outlines the methodology (including corpus, syntactic parser, and calculation). The results and discussion are presented in Section 3. Finally, concluding remarks are provided in Section 4.

DATA AND METHODS

Data

The Turkish oral data were drawn from the International Cross-Sectional Corpus of Japanese as a Second Language. We extract story-writing (1 task) and interaction (1 task) in Japanese by Turkish Japanese learners and native Japanese speakers (as a comparison). It should be noted that oral Japanese has three forms: plain, polite, and honorific/humble. The choice of spoken form depends on the social distance between

the utterer, the hearer, and the third person that is mentioned in the conversation. When the status of the hearer or the mentioned person is higher than the utterer, the honorification and humble form is used. When narrating a story, the plain or polite form is likely used. The polite form is rendered by a fixed form, that is, masu-form for verbs and desu-form for nouns and na-adjectives. The honorific and humble forms have two variations: (a) set expressions, and (b) for lexicons that do not have set expressions, honorification is realised via conjugation (prefix or suffix). The data details, including the amount, form, and content, are presented in Table 1.

Table 1. Study data

Material	Sample number	Total words	Content	Form
Story-writing	50	87300	The learner narrates a story based on the illustrations of the 4 and 5 columns	plain form, honorific and humble form
Interaction	50	24000	A natural conversation between learner and researcher is carried out for about 30 minutes	plain form, honorific and humble form

Analysis

The central goal of this study is to uncover how Turkish Japanese learners perform in oral tasks. To this end, lexical richness and syntactic complexity of Turkish Japanese learners' oral text are collected and measured. The MAMR and MAMSP are calculated for measuring lexicon sophistication. The MDD is employed for testing syntactic diversities. Additionally, the MAMR, MAMSP, and MDD are computed using self-written computer programme scripts.

Oral proficiency at lexical level

Cech and Kubat (2018), Covington and McFall (2010), Yan and Liu (2021), and Li, Liu, and Li (2022) confirmed that the moving window can obtain a better average type-token ratio (TTR). In light of previous work, the moving window of TTR in terms of word form is obtained via the following formula:

$$MATTR\ \ (W)_{word\ \ form} = \frac{\sum_{I=1}^{N-W+1} F_i}{\mathrm{W}\ \ (\mathrm{N}-\mathrm{W}+1)}$$

Obtain the moving window of TTR in terms of the lemma in the following formula:

$$MATTR\ (W)_{lema} = \frac{\sum_{I=1}^{N-W+1} F_i}{\mathrm{W}\ (\mathrm{N}-\mathrm{W}+1)}$$

Building on this, we can obtain lexical sophistication via $\frac{\sum_{I=1}^{N-W+1} F_i}{\mathrm{W}\ (\mathrm{N}-\mathrm{W}+1)}$ – $\frac{\sum_{I=1}^{N-W+1} F_i}{\mathrm{W}\ (\mathrm{N}-\mathrm{W}+1)}$

The higher the MAMR and MAMSP, the greater the lexical richness.

Oral proficiency at a syntactic level

Dependency distance (DD) is a concept under the Dependency Grammar framework (Tesnière 1959; Yngve 1960; Hudson 2007; Liu 2009b). It refers to the distance between the governor and the dependent. The governor acts as the core linguistic element in a sentence, that is, verb, predicate. The dependent is the subject, object, oblique, adverb, post/prepositional phrase, and so on. The distance between the governor and the dependent can be obtained via |governor − dependent| (Liu, Hudson, and Feng 2009). The MDD of the whole sentence would be:

$$\mathrm{MDD} = \frac{1}{\mathrm{n}}\sum_{i=1}^{n} |\mathrm{DD}_i|$$

Regarding the Japanese example (4), that is, 述べさせられましたか (nobe sase rare mashi ta ka), there are two types, five tokens of dependency relationships, for example, auxiliary and mark. The governor is the verb stem *nobe* 'speak'. *Sase, rare, mashi, ta, ka* are the dependents. The MDD is 2.5.

Table 2. Dependency relation and direction of sentence (4)

Dependency direction	Dependency relation
させ ← 述べ	aux
られ ← 述べ	aux
まし ← 述べ	aux
た ← 述べ	aux
か ← 述べ	mark

Turning to syntactic issues, we focus on word order. In Japanese, there are up to six ways of encoding events into linguistic expressions, under the condition that the verb, that is, the governor, appears at the end.

Variations of Japanese word order

AがBにCをV；AがCをBにV；BにAがCをV；

BにCをAがV；CをAがBにV；CをBにAがV

The basic word order of Turkish is SOV. However, the order is flexible; that is, six variations are possible: SOV, SVO, OSV, OVS, VSO, and VOS. Crucially, these variations differ from Japanese word order in that Turkish does not require the verb to be kept the end of the clause or sentence. Against this background, we are interested in how Turkish Japanese learners perform in a conversation or storytelling task in terms of word order; does the mother tongue play an essential role?

Calculation

MAMR, MAMSP and MDD and are calculated in the following steps:
Step 1: Draw raw data from the corpora
Step 2: Parse each sentence via the GiNZA v4 Parser (National Institute for Japanese Language and Linguistics, and Megagon Labs)
Step 3: Calculate the MAMR, MAMSP and MDD from the parsed outputs
Step 4: Produce a computer programme to proceed to statistical analysis

RESULTS AND DISCUSSION

Drawing on the methodology highlighted above, this section discusses Turkish Japanese learners' oral proficiency in storytelling and interaction (conversation). The lexical diversities calculated via the MAMR and MAMSP are outlined in Section 3.1. The syntactic diversity with the MDD as the metric is addressed in Section 3.2. Whether oral quality reflected by the MDD and their frequencies exhibits a specific regularity is explored in Section 3.2.

Lexical complexity of oral Japanese proficiency

The comparative data derived from native Japanese and Turkish Japanese learners are presented in Table 3.

Table 3. Turkish Japanese learners' oral proficiency (measured by MAMR and MAMSP): a comparison with native Japanese

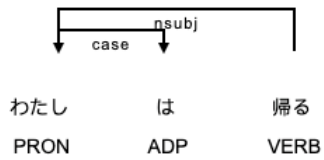
Turkish Japanese leaners	**Average MAMR**	**Average MAMSP**	**Native Japanese**	**Average MAMR**	**Average MAMSP**
Storytelling	0.013	1.029	Storytelling	0.0258	1.073
Interaction	0.011	1.032	Interaction	0.0133	1.037

In both storytelling and interaction, the data for Turkish oral Japanese texts indicate a close but smaller MAMR and MAMSP compared with that for native Japanese. From this, we can deduce that the L1 mother tongue does affect later language acquisition in terms of lexical sophistication.

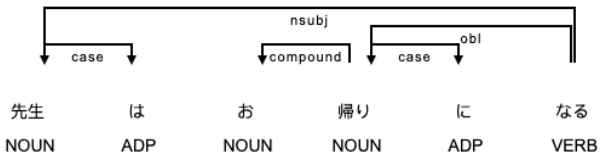
Syntactic diversity of oral Japanese proficiency

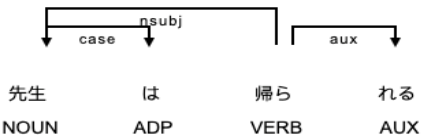
Having explored Turkish Japanese learners' oral proficiency at a lexical level, we now investigate syntactic mutuality. MDD is used as the metric. According to the results, the MDD of storytelling by Turkish Japanese learners ranges from 0.5 to 5.4. Most tokens reach a distance of 2.4. The average DD is 3.28. The MDD of storytelling by native Japanese ranges from 2.0 to 5.0. Most tokens go to the distance 2.2. The average DD is 3.87, slightly longer than storytelling by Turkish Japanese learners. The MDD of interaction (conversation) of Hungarian Japanese learners ranges from 0.6 to 5.8, indicating greater freedom. Most tokens go to the distance 1.0. The average DD is 3.18. The MDD of interaction by native Japanese ranges from 1.2 to 5.6. Most tokens go to the distance 1.2. The average DD is 4.42, longer than the interaction between Turkish Japanese learners and native Japanese-produced storytelling. This is due to the fact that interaction usually uses honorification and humble form whilst narrative (storytelling) usually uses the plain form. Recall that there are three levels of politeness in Japanese: casual, polite, and honorific/humble. The casual is presented in the plain form, and the polite is rendered by a fixed form, that is, *masu*-form for verbs and *desu*-form. The honorific and humble forms are realised either via fixed expressions or via prefix or suffix. Comparing (7)–(9), we derive a clearer understanding of the contention that the honorification and humble form lengthens word length and, in turn, sentence length.

(7) Watashi wa kaeru (plain form)
I TOP go back

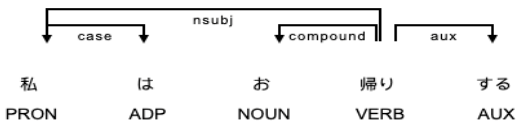


(8) Sensee wa okaerininaru/kaerareru (honorification form)
Teacher TOP go back





(9) Watakushi wa okaerisuru (humble form)
I TOP go back



The aforementioned findings are summarised in Table 4.

Table 4. Mean dependency distance of oral data for Turkish Japanese learners and native Japanese

	Storytelling			**Interaction**		
	minimal MDD	maximal MDD	the DD that attributes the most tokens	minimal MDD	maximal MDD	the DD that attributes the most tokens
Native Japanese	2.0	5.0	2.2	1.2	5.6	1.2
Turkish Japanese leaners	0.5	5.4	2.4	0.6	5.8	1.0

Compared to the oral data from native Japanese, Turkish Japanese learners' storytelling and interaction are similar regarding the freedom of MDD, the maximal MDD, and the DD that bears the most tokens. However, there is a distinction in the minimal MDD in both storytelling and interaction: oral data from Turkish Japanese learners is 0.5 compared to 2.0 for native Japanese. Moreover, the DD that attributes the most tokens in interaction is nearly half as short as that in storytelling. According to existing findings regarding simultaneous translation, the reason for this resides in the Principle of Least Effort (Zipf 1949), in that human action tends to try to lighten the processing load as much as possible.

Probability distribution of Turkish Japanese learners' oral proficiency

Having compared lexical sophistication and syntactic complexity of Turkish Japanese learners and native Japanese oral data based on a mathematical linguistic approach, the interaction between the two is found to be closer than for storytelling. We, therefore, conduct a further investigation of the regularity in the distributions and frequencies of the MDD. The curves in Figure 1 indicate the relationships between the DD, and their frequencies are concave downward. As shown in Table 5, three distribution models are fitted, that is, Harris (m, M), Dacey 3 (m, n), and Kelly (n, T), with 0.9515 and 0.9286 as the lowest and highest determination coefficient R^2, respectively ($R^2 > 0.90$, very good; $R^2 > 0.80$, good; $R^2 > 0.75$, acceptable; $R^2 < 0.75$, unacceptable).

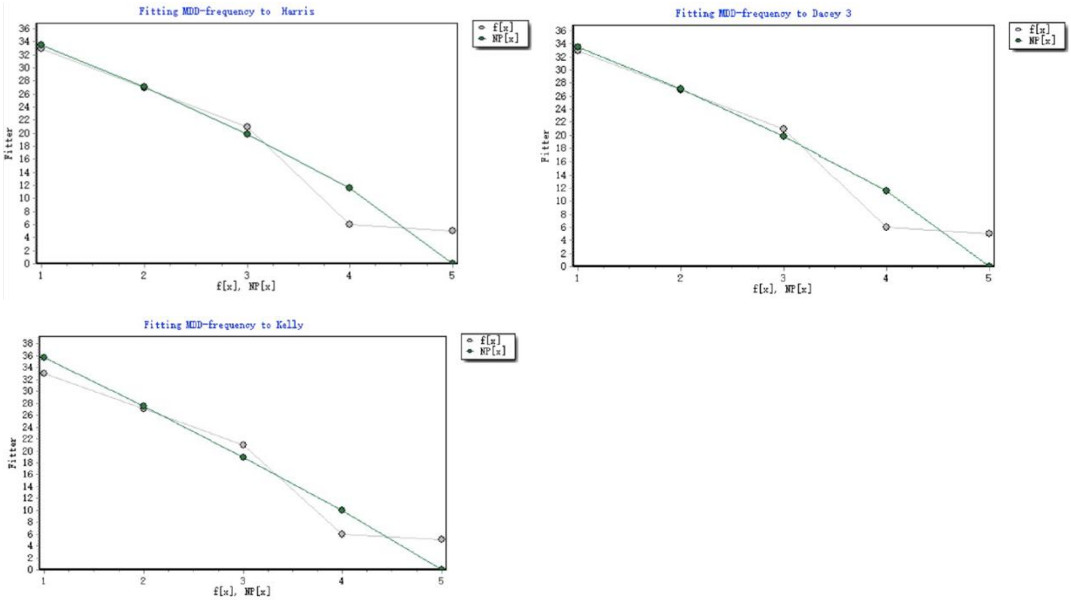


Figure 1. Fitting outcome of the relationship between the distribution of the MDD and their frequencies

Table 5. Fitting the distribution of the MDD and their frequencies to models

Fitting models	**Parameters**			
Harris (m, M)	m	M	X^2	R^2
	3.0000	1.7190	0.0983	0.9215
Dacey 3 (m, n)	m	n	X^2	R^2
	2.7183	3.0000	0.0982	0.9215
Kelly (n, T)	n	T	X^2	R^2
	4.0000	1.9017	0.5515	0.9286

CONCLUSION

This study incorporates a mathematical linguistic approach to examine Turkish Japanese learners' oral proficiency. Given the similarities between Turkish and Japanese, that is, morphologically agglutinative and syntactically SOV, it aims to understand whether language closeness influences language acquisition. Data were extracted from 150 hours of recordings of storytelling and 150 hours of interaction in Japanese by Turkish speakers, covering the plain, humble, and honorific forms. Oral proficiency is measured at lexical and syntactic levels. In terms of lexical sophistication, the MAMR and MAMSP are employed as the metrics. In terms of syntactic complexity, MDD is calculated.

Turkish Japanese learners' storytelling and interaction present a similar picture regarding the freedom of MAMR, MAMSP, MDD, the maximal MDD, and the DD that bears the most tokens to the oral data from native Japanese. We, therefore, contend that the L1 mother tongue affects later language acquisition morpho-syntactically. Moreover, the DD that attributes the most tokens in interaction is nearly half that of the DD in storytelling, confirming the Principle of Least Effort (Zipf 1949) that human action tends to lighten the processing load.

A further investigation of the regularity in the distributions and frequencies of MDD reveals that the MDD-frequency relationship in the Turkish oral data demonstrates a good fitting result to three models, that is, Harris (m, M), Dacey 3 (m, n) and Kelly (n, T).

Reference

Falhive, D., and. Snow, B. G. 1980. The use of objective measures of syntactic complexity in the evaluation of compositions by EFL students. In J. Oller & K. Perkins (Eds.), Research in language testing, pp. 171–176. Rowley, MA: Newbury House.

Harrington, M. 1986. The T-unit as a measure of JSL oral proficiency. Descriptive and Applied Linguistics19, 49–56.

Hunt, K.W. 1970. Syntactic maturity in school children and adults. Monographs of the Society for Research in Child Development, 35 (Serial No. 134).

Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A Research synthesis of college-level L2 writing. Applied Linguistics 4, 492–518.

Richards, B. 1987. Type/token ratios: What do they really tell us? Journal of Child Language, 14, 201–209.

Witte, S. P., and Sodowsky, R. E. 1978. Syntactic maturity in the writing of college freshmen.Paper presented at the Conference on College Composition and Communication. Denver, CO. (ERIC Document Reproduction Service No. ED 163 460)

Zipf, G. K. 1949. *Human behavior and the principle of least effort*. Cambridge: Addison Wesley Press.