

# Corpus of learners of Japanese as a L2 from 16 countries

Kumiko Sakoda

National Institute for  
Japanese Language and Linguistics



# Aims:

1. To introduce the large-scale learners' corpus of Japanese currently under construction
2. To analyze part of the data and discuss issues of grammar acquisition

# 1. Learners' Corpora of Japanese Language

Table 1. List of Learners' corpora (spoken)

Corpus Name	Data	Learners' Native Languages	FS	Level Check
KY Corpus	90 (30min.)	Chinese, Korean, English	×	OPI
KAIWA-DB (Cross-Sectional)	339 (30min.)	Chinese, Korean, English Indonesians, others	○	OPI
KAIWA-DB (Longitudinal)	About 20 46 dialogues (30min.)	Tagalog, Korean, Chinese, Russian, Malay, Portuguese	×	OPI

※ OPI : Oral Proficiency Interview (ACTFL)

**Table 1. List of Learners' corpora (spoken) cont.**

Corpus Name	Data	Learners' Native Languages	FS	Level Check
LARP	37 (20min) 3.5 years	Chinese	○	SPOT
BTSJ	294 dialogues 66 hours	Korean, Chinese, French	×	×
HATSUWA-TAISHO DB	190	Chinese, Korean, Thai, Japanese	○	SPOT (一部)
C-JAS	6 47 dialogues (60min) 3 years	Chinese, Korean	△	×

※ SPOT: Simple Proficiency Oriented Test

# Issues with corpora of learners of Japanese

1. Low number of learners
2. Most corpora contain data from English, Chinese or Korean native speakers; data for other languages is absent
3. Level of Japanese language proficiency is unclear
4. Background learner information is unavailable

## 2. The learner corpus under construction

### I-JAS

International corpus of Japanese  
As a Second language

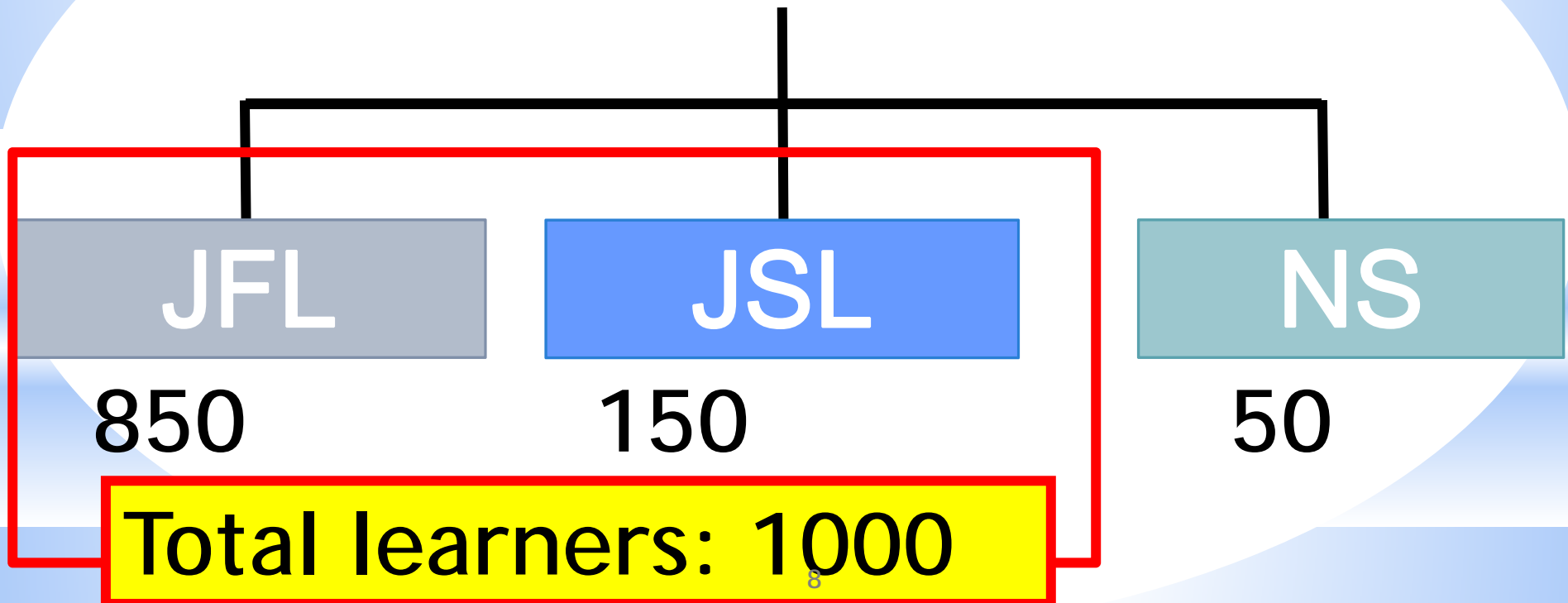
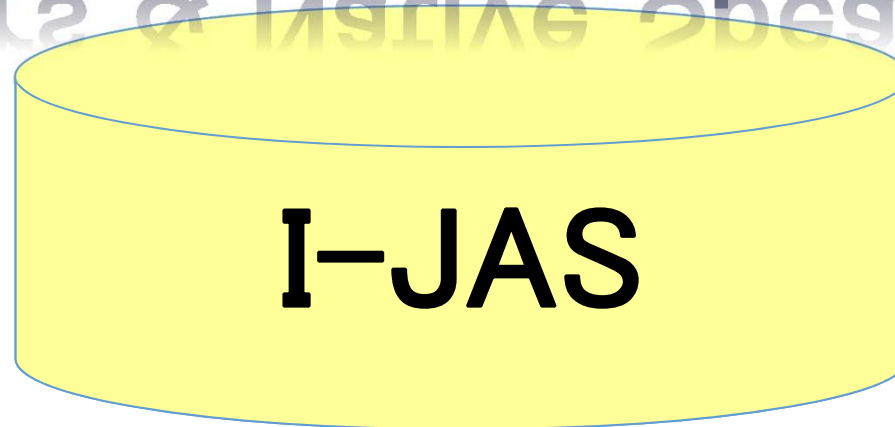
#### 【Aim】

To elucidate the effects on the acquisition process of different language environments, including differences in mother tongue

# Characteristics of I-JAS

- ① **Number of learners** and their breakdown
  - JFL Learners from 16 countries, speaking 12 native languages
  - JSL Class-room/ Natural Setting
  - Native Speakers
- ② **Detailed** background information
- ③ Objective Japanese Proficiency Tests (2 types)
- ④ A variety of tasks (6 types)
- ⑤ Release of text and **audio**

# Learners & Native Speakers





# Learners of JFL

**Chinese**  
200 learners

**German**  
50 learners

**French**  
50 learners

**Korean**  
100 learners

**Turkish**  
50 learners

**Spanish**  
50 learners

**English**  
100 learners

**Indonesian**  
50 learners

**Russian**  
50 learners

**Thai**  
50 learners

**Vietnamese**  
50 learners

**Hungarian**  
50 learners

# Learners of JSL & NS

Learners  
in  
Classroom  
Settings

100

Learners  
in  
Natural  
Settings

50

Native  
Speakers

20s

30s

40s

50

# Content (speech)

## 1. Story Telling

Look at 4–5 pictures and tell the story

## 2. Dialogue (30 minutes)

Semi-structured interview

The previous day's schedule/interest in Japan/home town/childhood memories/future/opinions etc.

# Content (speech)

## **3 . Role-play**

“Refusal” and “request” tasks

## **4 . Picture portrayal task**

Look at and describe in Japanese a single image

## **5 . Writing**

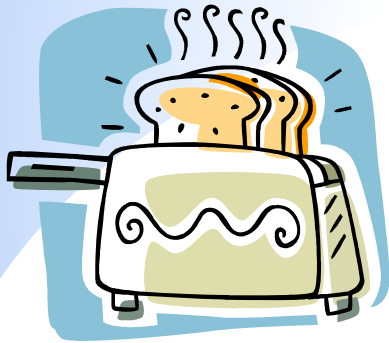
Look at the pictures used in 1. and write the story



Xu(2000)

# Content (composition)

## 1. Essay



“Our diets: fast food and home cooking”  
(around 600 characters)

## 2. Email

Establish 3 scenarios,  
then write emails  
(request, refusal etc.)



# Assessments of Japanese language proficiency

## 1. SPOT

(Simple Performance-Oriented Test)

Proficiency measured by testing aural comprehension

## 2. J-CAT

(Japanese Computerized Adaptive Test)

Computer-based proficiency test with automatic assessment

# In progress...

- ◆ Additional surveys abroad
- ◆ Ongoing learner surveys within Japan
- ◆ **Transcription of speech data**

**First release of data  
(Spring 2016)**

Second release  
(2017–Spring 2020)



# 3. Language use between different tasks

## Thai Speaker (TTH27)

### ◆ Speaking

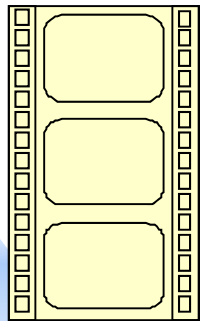
バスケットに入ったサンドイッチや果物、は犬を食べてしまいました

### ◆ Writing

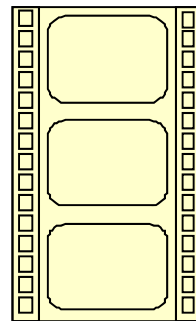
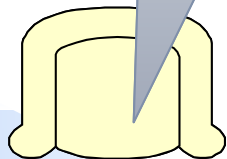
バスケットの中に入れられたサンドイッチやリンゴは犬に食べられてしまいました

## ■ Aims of the survey

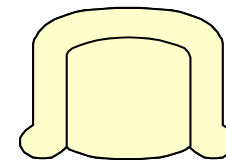
Using the same 5 images in “speaking” and “writing” tasks, is verb conjugation more accurate in the “writing” task?



Speaking  
task



Writing task



## ■ Learners

Chinese 15      English 15  
German 15      Thai 15

Table 2. Results of the Proficiency Tests

	Chinese	English	German	Thai
J-CAT	213	212	213	211
SPOT	69	68	68	68

# Story-telling (Picnic)

## 【Speaking】

Learner begins speaking as soon as they understand the content

- Conversation
- Role Play
- Description Task

## 【Writing】

Looking at pictures, the learner takes their time to write about the story



①

②

③

④

⑤

# Results of analysis (passive : ~しまう)

## (Okuno 2015)

**Table 3. Sentence Variations in Speaking and Writing**

	Ch. speaking	Ch. writing	En. speaking	En. writing	Ger. speaking	Ger. writing	Th. speaking	Th. writing
食べられて しまった	1	0	1	4	2	2	2	7
食べられた	6	10	1	0	1	1	3	1
食べて しまった	1	2	9	7	7	7	5	6
食べた	5	2	4	2	3	3	4	1
食べかけた	0	0	0	0	1	0	0	0
なくなった	0	0	0	2	1	1	0	0
食べ切った	0	0	0	0	0	1	0	0
叙述なし	2	1	0	0	0	0	1	0

● From Table 3.

	Ch. speaking	Ch. writing	En. speaking	En. writing	Th. speaking	Th. writing
食べられて しまった	1	0	1	4	2	7
食べられた	6	10	1	0	3	1
食べてしまった	1	2	9	7	5	6

- Errors by Chinese, English and Thai learners involving the passive constructions 「食べられた」 and 「食べられてしまった」 were more prevalent in the writing task than the speaking task
- German learners were the exception

# Change in forms used by the same learner

**Table 4. Variations among Speaking and Writing by the same learner**

	Speaking	Writing		Speaking	Writing
C23	食べて	食べられました	T10	残りました	食べてしまいました
C28	食べました	食べられました	T18	食べられました	食べられてしまいました
C49	食べました	食べてしまいました	T19	全部食べました	食べられてしまいました
C52	食べられちゃった	食べられました	T22	食べられてしまいました	食べてしまいました

● From Table 4

1. The **written task** data contains more instances of use of the passive and the 「てしまふ」form



## 2. We can suppose the following acquisition process

**(passive)**

食べた  
食べました

食べられた  
食べられました

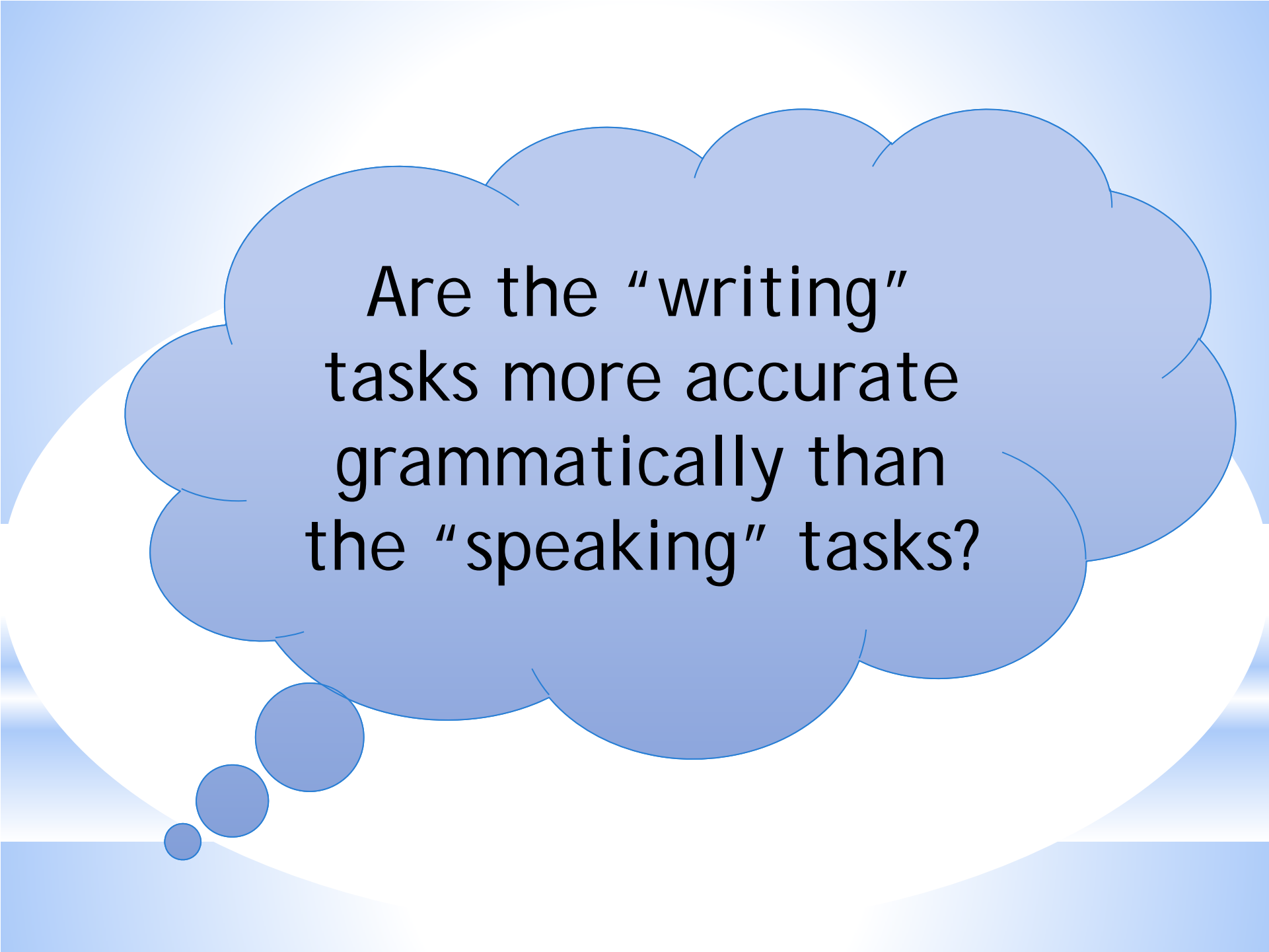
食べられてしまいました

食べた  
食べました

食べてしまった  
食べてしまいました

食べられてしまいました

**【～てしまう】**



Are the “writing”  
tasks more accurate  
grammatically than  
the “speaking” tasks?

# Results of analysis (Sakoda 2014)

Table 5. Intransitive and Transitive Verbs

Student	Speaking Task	Writing Task	Intrans. / Trans. Verb
C37	SDをBKに <b>入れた</b> (誤) 犬がBKに入った(正)	SDをBKに <b>入る</b> 時(誤) 犬はBKに入って(正)	<b>入る</b> X <b>入れる</b>
E27	BKを <b>あいた</b> 後で犬が (誤)	BKを <b>あいた</b> ところ犬が (誤)	<b>あく</b> X <b>あける</b>
G21	BKの中をSDに <b>入りま</b> した(誤)	犬が入りました(正)	<b>入る</b> X <b>入れる</b>
T49	BKを <b>あく</b> と(誤)	BKを <b>あく</b> と(誤)	<b>あく</b> X <b>あける</b>
	S D サンドイッチ	B K バスケット	

## ● From Table 5

There was no change observed in the use of intransitive and transitive verbs (both tasks showed the same usage trends)



For transitive–intransitive verb pairs, a tendency was observed to favor use of one or other of the pair

**Item  
learning ?**

## 4. Conclusions

What this study revealed:

(1) Using the same images to conduct “speaking” and “writing” tasks with the same learners, there were areas where differences were observed and those where none was observed.



Differences in the tasks (thinking time) may or may not have an effect

(2) There was a trend for passives and the 「～てしまう」 construction to be used when writing, even if they were not used in the speaking task

食べた



食べました

食べ**られ**た



食べ**てしま**った

食べ**られてしま**いました

Learners have sufficient time to use correctly grammatical structures that they have studied

**System  
learning ?**

(3) The trend is for there to be no change in the use of intransitive–transitive verb pairs between spoken and written language.

Intransitive and transitive verbs may be being processed as lexical rather than grammatical items

**Item  
learning ?**

# What we can discover from learner corpora



Learners' use of Japanese  
||  
A partial view of learners'  
grammar

**Native  
Language**

**Learning  
Environment**

**Task  
Variation**



## Sources

1. 奥野由紀子(2015)「「話す」課題と「書く」課題に見られる中間言語変異性—ストーリー—描写課題における「食べられてしまった」部を対象に—」2014年度国立国語研究所共同研究発表会予稿集 20-23.
2. 迫田久美子(2014)「書き言葉と話し言葉の違い—学習者コーパスに見る言語運用—」ICPLJ2014, パネルセッション,
3. 許夏珮(2000)「自然発話における日本語学習者による『テイル』の習得研究—OPIデータの分析結果から—」『日本語教育』104, pp.20-29

Thank you for your attention.



Kumiko Sakoda

