

The 3rd International Workshop on Advanced Learning Sciences

01/08/2015

Comparative Analysis of English, Japanese and Chinese Based on Cross-Linguistic Learner Corpora

Keiko Mochizuki, Sano Hiroshi, Caroline Kano,
Laurence Newbery-Payton



Structure of the Presentation

Introduction of the corpus

Statistical testing

Error analysis

Future avenues for research

Overview of the “Sunrise Corpus” # (Kaken B)

Sub-corpus	Total files	Total words
TUFS*	452	202448
SISU**	62	17919
Taiwan***	43	19813
Total	557	240180

As of 31/05/2015

* Tokyo University of Foreign Studies

** Shanghai International Studies University

*** National Normal Taiwan University and other
Taiwanese universities

Original Essay

TUFS_01_2013↵

↵

Culture of “Entertainment”: China, England and Taiwan (1)↵

↵

At Tokyo University of Foreign Studies (TUFS), I teach not only Japanese students but also exchange students from Taiwan University, the University of Leeds, and the School of Foreign Languages, Peking University. When the students from these universities get accustomed to their new lives after coming to Japan, I always make it a rule to invite them to my house. ↵

When I serve a meal the contents of it are always sushi, dim sum steamed in a steaming basket, sandwiches, pumpkin salad, assorted fruits and grape juice. It is true that they are all simple because I can make them just by assorting what I bought at the store. In fact, they can be called “global food” because the students from all over the world can eat them regardless of their taste in food. Above all, I think the vivid impression of having sat around the table at my house, surpassing that of having a meal at a luxurious restaurant, would be a precious memory of studying abroad.↵

Revised Essay

TUFS_01_2013+

Culture of "Entertainment": China, England and Taiwan (1)+

At Tokyo University of Foreign Studies (TUFS), I teach not only Japanese students but also exchange students from Taiwan University, the University of Leeds, and the School of Foreign Languages, Peking University. When the students from these universities get accustomed to their (new) lives after coming to Japan, I always make it a rule to invite them to my house. +

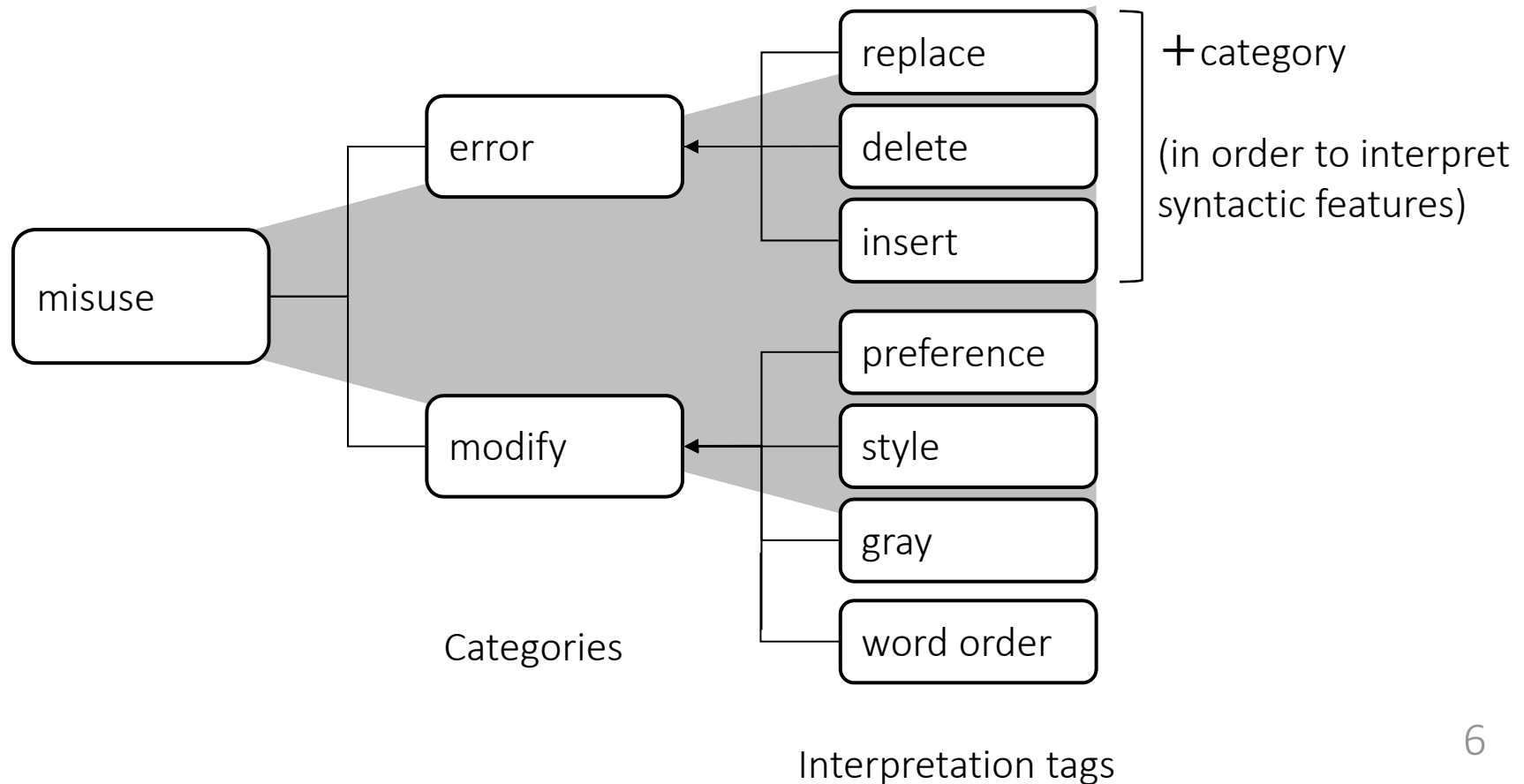
When I serve (a meal) the contents of it are always sushi, dim sum steamed in a steaming basket, sandwiches, pumpkin salad, assorted fruits and grape juice. It is true that (they are all) simple because I can make them just by assorting what I bought at the store. In fact, they can be called [global food] because the students from all over the world can eat them regardless of their taste (in food). Above all, I think the vivid impression of having sat around the table at my house, (surpassing) that of having a meal at a luxurious restaurant, would be a precious memory of studying abroad. +

During my 20s and 30s, I visited Beijing, Shanghai, London and Taiwan. A (particular) memory of when I studied abroad is of when my teachers invited me to their

- 作成者 Preference: 'Lives' is too ▼
- 作成者 Error: Have suggests a ▼
- 作成者 削除: with ▼
- 作成者 削除: have ▼
- 作成者 Preference - When suggest ▼
- 作成者 削除: Whey ...hen I serve ▼
- 作成者 Gray ... They are all is ▼
- 作成者 削除: each of them is ...(they ▼
- 作成者 Error- The Insertion of ▼
- 作成者 Preference - eat is a ▼
- 作成者 削除: have ...at them ▼
- 作成者 Agreement error. If ▼
- 作成者 削除: which ...urpassing) ▼
- 作成者 'studying abroad' is a ▼
- 作成者 削除: have ...isited Beijing ▼
- 作成者 Preference - Particular ▼
- 作成者 削除: I remember ⁵ s the ▼

Interpretation Framework (error tags)

Classification and in-text marking of syntactical, stylistic and rhetorical errors



Rules for Correction <error>

<p><replace></p>	<p>Using cell phones <u>is</u> are so easy.</p> <p>Write strike-through “are” on the error word(s) and change color.</p> <p>Write the correct word(s) by coloring function.</p>
<p><delete></p>	<p>They are often have a bad influence on us.</p> <p>Write strike-through “are” on the error word(s) and change color.</p>
<p><insert></p>	<p>Students have to pay <u>a</u> tuition fee.</p> <p>Write the correct word(s) by coloring function.</p>

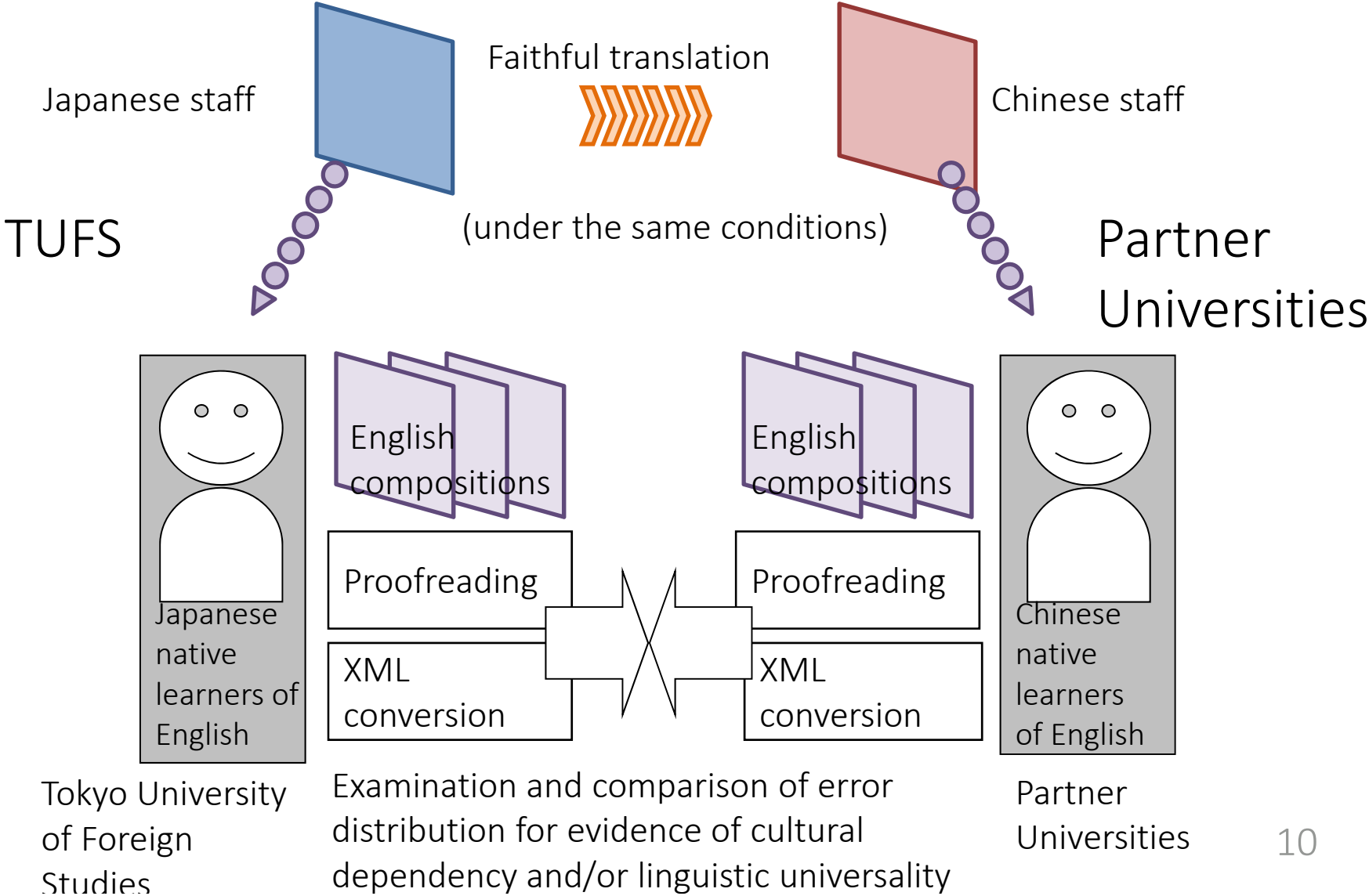
Rules for Correction <modify>

<p><preference></p>	<p><preference> means that the word(s) is not incorrect. However, the different word is better.</p> <hr/> <p>However, I do not think that education should be treated <u>(in the same way)</u> as other kinds of services.</p> <hr/> <p>After using “track change”, add parenthesis ().</p>
<p><style></p>	<p><style> means that the word(s) is not incorrect. However, the word is not used in the style of the essay. *Academic Writing</p> <hr/> <p>They <u>((do not))</u> don't pay a tuition fee.</p> <hr/> <p>After using “track change”, add double parenthesis (()).</p>
<p><gray></p>	<p><gray> means that the correction is difficult to classify into the other categories.</p> <p><gray> includes the case that you can't understand what the writer want to express.</p> <hr/> <p>I have had <u>only</u> a few opportunities to <u>(#spend my life living-)live</u> abroad and to speak English, so I don't think I'm a good English speaker.</p> <hr/> <p>After using “track change”, add parenthesis and hash (#).</p>

XML Annotation

```
<?xml version="1.0" encoding="UTF-8"?>
- <composition>
  - <body>
    <id>TUFS_01_2013</id>
    <task>Task1</task>
    - <learner_data>
      <toeic>TOEIC:820</toeic>
      <length_of_learning>英語学習歴:7</length_of_learning>
      <experience_of_living_overseas>海外滞在歴:なし</experience_of_living_overseas>
      <native_language>母語:日本語</native_language>
      <first_educational_language>第一教育言語:日本語</first_educational_language>
      <second_educational_language>第二教育言語:なし</second_educational_language>
    </learner_data>
    <title>Self-Introduction</title>
    - <paragraph>
      Hi! I'm ○○. Everybody calls me ○○-chan. I think it is easier to call my family name
      because '○○' is shorter than my given name. So please call me ○○. I was born and
      bred in Tokyo. I have only a few opportunities to spend my life living abroad and
      speaking English, so I don't think I'm a good English speaker.Hi! I'm ○○. Everybody
      calls me ○○-chan. I think it is easier to call
      <error category="category1" revised="" type="insert"> </error>
      my family name
      <error category="category1" revised="" type="insert"> </error>
      because '○○' is shorter than my given name. So please call me ○○. I was born and
      bred in Tokyo. I have
      <error category="category1" revised="" type="insert"> </error>
      only a few opportunities to
      <modify category="gray" revised="">spend my life living</modify>
      abroad and
      <modify category="" revised="">speaking</modify>
      English, so I don't think I'm a good English speaker.
    </paragraph>
```

Error Comparison



オンライン英作文学習者コーパス・誤用辞典

Online Dictionary of Misused English

———— Based on a Learners' Corpus ————

[▶ ホーム](#) | [▶ プロジェクトについて](#) | [▶ コーパスについて](#) | [▶ 利用方法](#)

○ Login


ログインID

パスワード

[ログイン](#)

[パスワード再発行](#)

○ ホーム

この英語学習者コーパスは、東京外国語大学国際日本研究センター  および英語専攻の支援のもとに、英作文教育研究を目的として公開する添削情報、誤用情報を付した英作文学習者コーパスです。

構築は、東京外国語大学総合国際学研究院望月圭子研究室、佐野洋研究室、英語専攻外国人教員キャロライン・狩野、マシュー・ミラーの連携で行われています。

画面右上の「新規登録」から利用申請をすれば、無料で閲覧することができます。

○ MENU

[ホーム](#)

- [▶ プロジェクトについて](#)
- [▶ コーパスについて](#)
- [▶ 利用方法](#)

○ Online Status

1
ログアウト

U
ホーム
プロジェクトについて
タスクについて
過去

(ODME)
タスクダウンロード

ie Status
メンバー ▶ 1人
メンバー ▶ 1人

誤用 → 訂正

※語の過不足を検索する場合は「φ」または「phi」と入力してください。

+ タスクから検索

+ 学習者情報から検索

◀ ◁ 1 - 10件表示 / (26件中) ▷ ▶ 表示件数 ページ数

◇	学習者ID	◇	原文	◇	訂正文	◇	誤用	◇	訂正	◇	詳細	◇
1	TUFS_10_2012		I was born in Nagano prefecture, and I lived there until the middle of my 3rd year in elementary school.		I was born in Nagano prefecture, and I lived there until the middle of my 3rd year of elementary school.		in		of		詳細	
2	TUFS_29_2012		The experience in this programme is very important to me.		The experience of this programme is very important to me.		in		of		詳細	
3	TUFS_22_2012		The strict atmosphere in a library might force you to finish the task.		The strict atmosphere of a library might force you to finish the task.		in		of		詳細	
4	TUFS_17_2012		Many people in urban areas long to live in the countryside in order to escape from the pressure and stress in the city.		Many people in urban areas long to live in the countryside in order to escape from the pressure and stress of the city.		in		of		詳細	
5	TUFS_17_2012		He said to me that the life in the country was too hard for him because he was accustomed to the convenient life in the city.		He said to me that the life in the country was too hard for him because he was accustomed to the convenient life of the city.		in		of		詳細	
6	TUFS_19_2012		Therefore, there is a big difference in the process in learning English between children		Therefore, there is a big difference in the process of learning English between children		in		of		詳細	

Translation Task

“Traditions of ‘Hospitality’ in China, Britain and Taiwan”

Japanese original → Chinese and English versions

Japanese version: 781 characters

Chinese version: 675 characters

Data Set	Number of Files	Number of Words
SISU	62	17919
TUFS	41	13119

Learners' English Level

TUFS:

Studying English for 8 ½ years on average
TOEIC: 745 points

SISU:

Studying English for 12 years on average
CET4* 580 points (15/62 learners)
CET6* 598 points (30/62 learners)

* CET4/CET6: College English Test,

Proficiency Test Conversion Table

(http://language.sakura.ne.jp/s/kaken_icnale.html)

	TOEFL(iBT)	TOEFL(PBT)	TOEIC	IELTS	Cambridge	STEP	TEPS	CET, TEM	VST
A2	*	*	225	*	KET	*	*	*	20
B1.1	57	487	550	4	PET	2	417	CET4	29
B1.2	72	533	650	4.5	*	*	502	*	32
B2.1	87	567	785	5	FCE	Pre1	608	TEM4/CET6	35
B2.2	99	597	860	6	*	*	700	*	38
C1	110	637	935	7	CAE	1	828	TEM8	41

Focus of the Study

Preposition	Instances of Overuse	Instances of Underuse	Total instances of misuse
At	66	155	221
In	189	92	281
Of	95	61	156
At/in/of (total)	350	308	658
Total	555	562	1117
At/in/of proportion of total (%)	63.1	54.8	58.9

Error Example: “overuse” and “underuse”

(1) “I would like to talk about my memory ~~in~~ *of* Shanghai first.”

→ Overuse of *in*

→ Underuse of *of*

Error Example: “overuse” and “underuse”

(2) “After I sat tight, teacher would put a handful of Longjing tea into a traditional Chinese teacup ~~of traditional Chinese flavor~~ with a top.”

→ Overuse of *of*

Chi squared results (all errors)

Preposition	Chi squared value (DF:1)	P value	Higher frequency corpus
at	16.71	0.0000	SISU
by	3.31	0.0688	
during	2.85	0.0915	
for	2.56	0.1093	
from	19.79	0.0000	TUFS
in	1.01	0.3156	
into	0.00	1.0000	
of	14.66	0.0001	TUFS
on	5.15	0.0233	TUFS
up	10.32	0.0013	TUFS

Colour coding

0.1%以下

1%以下

5%以下

石川慎一郎・前田忠彦・山崎誠編『言語研究のための統計入門』付属ディスク.2010.くろしお出版

Chi squared results (overuse of X)

Preposition	Chi squared value (DF:1)	P value	Higher frequency corpus
at	5.74	0.0166	TUFS
by	2.63	0.1045	
during	4.39	0.0361	TUFS
for	0.66	0.4150	
from	4.98	0.0256	TUFS
in	20.14	0.0000	SISU
into	0.04	0.8469	
of	11.56	0.0007	TUFS
on	0.22	0.6417	

(3) “He always wore a smile and offered me with the candies *in from* a bright red box just like the wedding candy box.” (SISU)

→ Overuse of *in*

Colour coding

0.1%以下

1%以下

5%以下

Chi squared results (underuse of X)

Preposition	Chi squared value (DF:1)	P value	Higher frequency corpus
at	42.30	0.0000	SISU
by	0.08	0.7840	
during	0.00	1.0000	
for	1.29	0.2562	
from	13.87	0.0002	TUFS
in	21.03	0.0000	TUFS
into	0.01	0.9158	
of	3.08	0.0791	
on	9.39	0.0022	TUFS
up	10.85	0.0010	TUFS

(4) “At that time, professors ~~in~~ *at* Fudan University had no rooms for research, so he lived in the accommodation which was next to the university.” (TUFS)

→ Underuse of *at*

Colour coding

0.1%以下

1%以下

5%以下

Errors involving *in/of*: overuse of *in*

Correct Use	SISU Freq.	TUFS Freq.	Higher freq. corpus	Log Score	Significance level
at	108	16	SISU	32.65	p < 0.0001
from	6	8	TUFS	6.14	p < 0.05
into	13	8	TUFS	1.53	—
∅	4	2	TUFS	0.17	—
of	4	9	TUFS	10.73	p < 0.01
on	5	5	TUFS	2.67	—
to	1	1	TUFS	0.53	—
Total	141	49			

(5) “I would like to talk about my memory *in* *of* Shanghai first.”
(SISU)

→ Overuse of *in*

Colour coding
0.1%以下
1%以下
5%以下

Errors involving *in/of*: underuse of *in*

Incorrect Use	SISU Freq.	TUFS Freq.	Higher freq. corpus	Log Score	Significance level
at	21	23	SISU	37.60	p < 0.0001
for	1	0	SISU	2.18	—
from	0	2	TUFS	1.64	—
∅	5	13	SISU	0.29	—
of	2	21	TUFS	8.02	p < 0.01
to	2	1	SISU	1.35	—
with	0	1	TUFS	0.82	—
Total	31	61			

(6) “I got a master’s degree ~~of the~~ *in* Chinese language at TUFS and from 1986 to 1988 I studied at Fudan University as a government-financed foreign student.” (TUFS)

→ Underuse of *in*

Colour coding

0.1%以下

1%以下

5%以下

Errors involving *in/of*: overuse of *of*

Correct Use	SISU Freq.	TUFS Freq.	Higher freq. corpus	Log Score	Significance level
as	0	1	TUFS	0.99	—
at	11	5	SISU	5.80	p < 0.05
for	1	6	TUFS	2.07	—
from	0	4	TUFS	3.95	p < 0.05
in	2	21	TUFS	10.91	p < 0.001
∅	20	12	SISU	7.22	p < 0.01
on	2	9	TUFS	2.22	p < 0.01
with	1	0	SISU	1.89	—
Total	37	58			

Colour coding
0.1%以下
1%以下
5%以下

(7) Then, he gave me a candy with a smile from a red candy box which is like a gift ~~of~~ *from* a wedding party.
 → Overuse of *of*

Errors involving *in/of*: underuse of *of*

Incorrect Use	SISU Freq.	TUFS Freq.	Higher freq. corpus	Log Score	Significance level
about	2	0	SISU	3.04	—
by	0	2	TUFS	2.52	—
during	1	8	TUFS	5.33	p < 0.05
for	0	1	TUFS	1.26	—
from	1	4	TUFS	1.56	—
in	4	9	TUFS	1.38	—
on	2	0	SISU	3.04	—
that	2	0	SISU	3.04	—
to	1	0	SISU	1.52	—
∅	15	6	SISU	5.24	p < 0.05
while	0	1	TUFS	1.26	—
with	1	2	TUFS	0.22	—
Total	29	33			

Colour coding

- 0.1%以下
- 1%以下
- 5%以下 24

Error Comparison

Error	SISU frequency	TUFS frequency	Total
“memory/ memories <i>in</i> ”	2	8	10
“professor <i>in</i> Chinese”	1	1	2
“savor filled <i>in</i> the Babao rice”	1	0	1
“bedroom <i>of</i> dormitory”	1	7	8
“menu <i>of</i> Chinese restaurant”	1	0	1
“degree <i>of</i> Chinese”	0	12	12
“life <i>of</i> those days”	0	1	1
“time <i>of</i> life”	0	1	1
Total	6	30	36

Comparison with original texts: overuse of *in*

Error	Chinese original	Japanese original
“memory/ memories <i>in</i> ”	(8a) 首先，就让我谈谈在上海留学时的一段回忆。	(8b) まず、最初に、上海留学中の思い出についてお話しします。
“professor <i>in</i> Chinese”	(9a) 我的指导教授是著名的汉语语言学家胡裕树教授。	(9b) 著名な中国語学者であった胡裕樹教授
“savor filled <i>in</i> the Babao rice”	(10a) 刚蒸好的八宝饭所带有的那种“软软、热热、甜甜”的幸福滋味，到现在仍然记忆犹新	(10b) 八宝飯の「やさしく、柔らかく、幸福な甘さ」

Analysis: “memory/memories *in*”

(11a) 「^{zai}在上海留学^{shi de}时的一段回忆」

上海で留学する時 の思い出

→ memories of [time [in Shanghai study]]

- 「中」 in Japanese: both spatial and temporal meanings
Not dependent on word order

(11b) 「上海留学中の思い出」

上海で留学する期間 の思い出

→ memories of [in period of [Shanghai study]]

Other instances of overuse of *in*

“professor in Chinese”

「汉语语言学家」 「中国語学者」



Compound nouns

Low error frequency, and no difference in error frequency

“savor filled *in* the Babao rice”

「八宝饭所带有的...滋味」

Perceiving property as spatial?

Comparison with original texts: overuse of *of*

Error	Chinese original	Japanese original
“bedroom <i>of</i> dormitory”	(12a) 在紧邻大学的老师 宿舍里 的 书房 兼寝室里进行的	(12b) ご 自宅の書斎 兼寝室で
“menu <i>of</i> Chinese restaurant”	(13a) 每当在 中国餐馆里 看到八宝饭	(13b) 中国料理店で 、八宝飯をみつけると
“degree <i>of</i> Chinese”	—	(14) 中国語学の修士号
“life <i>of</i> those days”	(15a) 虽然是一个物资不是很丰裕的时代	(15b) とても質素な時代でした
“time <i>of</i> life”	(16a) 在我二三十岁的 时候 也曾经到北京、上海、伦敦以及台湾留学过。	(16b) 私は、20代から30代にかけて、北京、上海、 ロンドン 、台湾に留学したことがあります。

Analysis: “bedroom *of* dormitory”

(12b) 「ご自宅の書斎兼寝室で」

→ lexical meaning + 「で」case marker → spatial meaning

Influence of 「の」

(12a) 「在紧邻大学的老师宿舍里的书房兼寝室里进行的」

→ 「里」^{li} required even with place nouns

Tendency to express spatial meaning using “in”

Analysis: “degree *of* Chinese”

(14) 「中国語学**の**修士号」

Influence of 「の」

(9b) 「著名な**中国語学者**であった胡裕樹教授」の文と対照的

Just 3 out of 30 errors by Japanese learners appear to be unrelated in any way to 「の」

Further avenues of research

Comprehensive analysis of all の errors

de
Comparison of Of ▪ の ▪ 的 in English, Japanese and Chinese

Errors between *at* and *in*

Spatial awareness

As there was no telephone ~~at~~ in his dormitory, I used to pay him an unexpected visit.

Temporal awareness

I studied in Beijing, Shanghai, London, and Taiwan ~~at~~ during/in my twenties and thirties.

Sources / Resources

Corpus of Contemporary American English

<http://corpus.byu.edu/coca/>

(2015年7月10日参考)

石川慎一郎・前田忠彦・山崎誠編『言語研究のための統計入門』付属ディスク.2010.くろしお出版

石川慎一郎研究室ホームページ

http://language.sakura.ne.jp/s/kaken_icnale.html

(2015年7月4日参考)

Lancaster University Log Likelihood Calculator <http://ucrel.lancs.ac.uk/llwizard.html>

(2015年6月6日参考)

望月圭子・キャロライン狩野(2005)「英語・日本語における空間・時間に関わる格標識:日本語母語話者による英作文学習者コーパスにみられる誤用類型」『東京外国語大学論集』85:219-236

Online Dictionary of Misused English – Based on a Learners' Corpus

http://sano.tufs.ac.jp/lcshare/htdocs/?action=pages_view_main&page_id=49

(2015年6月8日参考)

Rimmington, Don and Po-Ching Yip (2004) Chinese A Comprehensive Grammar. London: Routledge.