

The Construction of A Chinese Language Error Corpus

Yaming SHEN, ZHANG Zheng, Keiko MOCHIZUKI



1. Objectives of Constructing the Learners' Error Corpus

- (1) Discover the grammatical errors made by advanced-level learners
 - reflect significant differences in conceptual representation between the target language and the native language
- (2) Obtain new findings for comparative linguistics
 - language-specific error types based on the learners' native language
 - universal error types which occur regardless of the learners' native languages
- →Improve language teaching methodologies

2. Data Sources

	TUFS	NTNU
Number of learners	165	Over 600
Mother tongue	Japanese	English
Learner data	College students in their 2 nd , 3 rd , or 4 th year	Students taking TOCFL*
Learner level (CEFR** equivalent)	A2-B2	A2-B2

*TOCFL: The Test of Chinese as Foreign Language

**CEFR: The Common European Framework of Reference for Languages

2. Data Sources

Example of Learner' s Profile (TUFS)

1	Learner's ID	Th_Ch_001	13	Study Abroad Experience Institution / Period	Mandarin Center, National Taiwan Normal University, August1-31. 2014
2	Name	Tokyo Taro	14	Speaking with my family	Japanese
3	Major	Chinese	15	Speaking with friends	Japanese
4	Year	3	16	Language used in Elementary School	5-9 English, 9-12 Japanese
5	Gender	male	17	Language used in Junior High School	Japanese, English
6	Age	21	18	Language used in Senior High School	Japanese, English
7	Nationality	Japan	19	Test of Chinese as a Foreign Language (TOCFL)	Band B(2014)
8	Residential History	Canada 4-9 ; Japan 0-4,9-21	20	HSK 汉语水平考试	5 級 (2012)
9	Native Language	Japanese	21	English TOEFL(iBT)	108 (2013)
10	Language of Education	Japanese, English	22	TOEIC	955 (2012)
11	Length of Chinese study	3 years and 2 months	23	IELTS (academic)	8.0 (2013)
12	Institution	Tokyo University of Foreign Studies			

2. Data Sources

	TUFS	NTNU
Type of data	Composition as take-home task	Composition as test
Size of data	216,727 characters	237,475 characters
Number of compositions	471	691

3. Data Processing

- 1. Correct compositions
- 2. Add tags to instances of misuse
- 3. Convert files into XML format
- 4. Upload to online research platform

東京外国語大学 国際日本語センター・国際日本語教育研究センター(2010-2015年度)
科学研究費助成事業 基盤研究(6)「高度中国語コミュニケーション構築と母語文化の高度・日本語-中国語教育実践」

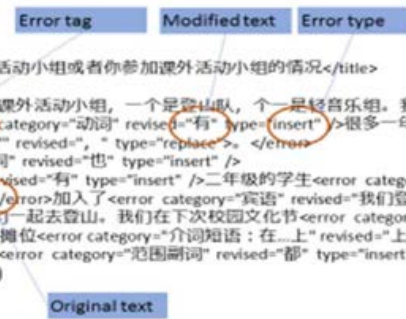
英語学習者 誤用検索サイト -en → in	日本語学習者 誤用検索サイト は → が	中国語学習者 誤用検索サイト 上 → 里
Learners' Error Corpora of Chinese Searching Platform	Learners' Error Corpora of English Searching Platform	Learners' Error Corpora of Japanese Searching Platform
上 → 里	-en → in	は → が
日语学習者作文 誤用検索プラットフォーム は → が	汉语学習者作文 誤用検索プラットフォーム 上 → 里	英語学習者作文 誤用検索プラットフォーム -en → in

© Copyright 2016, 東京外国語大学. All Rights Reserved.



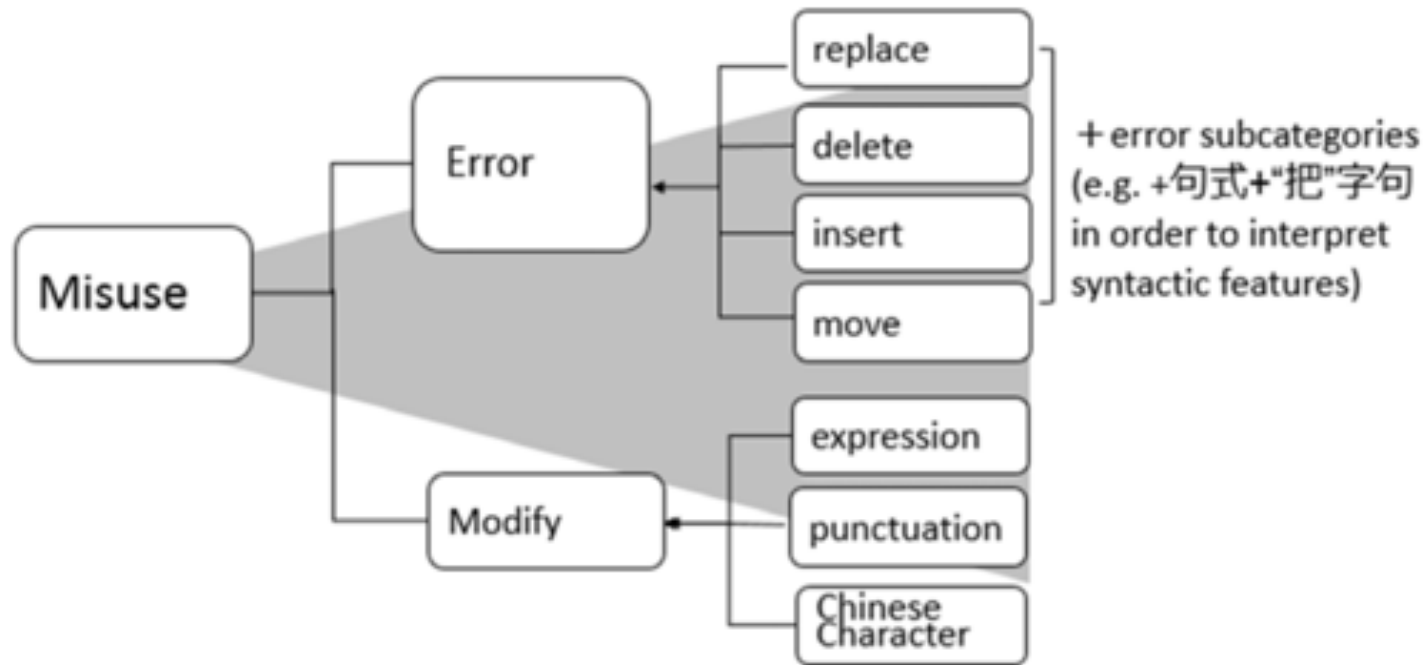
Digitization Framework (XML Data)

```
<?xml version="1.0" encoding="UTF-8"?>
<composition>
  <id>Tu_Ch_021</id>
  <learner_data>
    (... snip ...)
  </learner_data>
  <title>介绍一下大学的课外活动小组或者你参加课外活动小组的情况</title>
  <body>
    <paragraph>我参加了两个课外活动小组，一个是登山队，一个是轻音乐组。我是登山队的副队长，负责会计。今年<error category="动词" revised="有" type="insert" />很多一年级的学生加入了我们登山队<error category="" revised="", type="replace" />。</error>
    <error category="语气副词" revised="也" type="insert" />
    <error category="动词" revised="有" type="insert" />二年级的学生<error category="语气副词" revised="" type="delete" />也<error>加入了<error category="宾语" revised="我们登山队" type="insert" />我期待跟他们一起去登山。我们在下次校园文化节<error category="结构助词" revised="的" type="insert" />摊位<error category="介词短语；在...上" revised="上" type="insert" />，要卖火锅。我们的火锅每年<error category="范围副词" revised="都" type="insert" />很受欢迎。我们登山队也卖烟。 (... snip ...)
  </paragraph>
</body>
</composition>
```



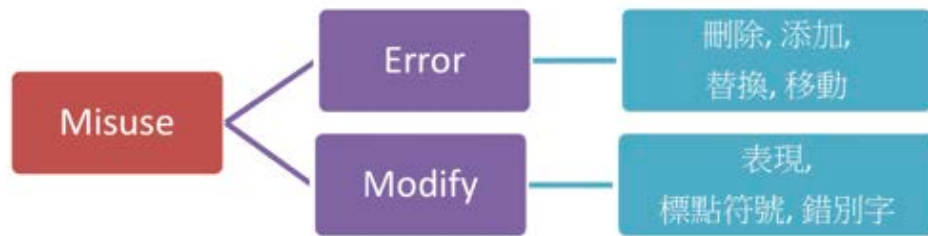
4. Error Tag Categories

Interpretation Framework (error tags)
Classification and in-text marking of syntactical, lexical, stylistic, rhetorical and notational misuses



4. Error Tag Categories

A. Subcategories of Misuse



B. Subcategories of Error

	大分類	小分類
1	名詞	時間名詞, 處所名詞, 方位詞
2	數詞	
3	量詞	
4	動詞	狀態動詞, 動作動詞, 存現動詞, 關係動詞, 能願動詞, 趨向動詞, 使令動詞
		及物動詞, 不及物動詞, 雙賓動詞
		重疊動詞
5	形容詞	
6	副詞	程度副詞, 範圍副詞, 時間副詞, 情態副詞, 否定副詞, 語氣副詞, 關聯副詞
7	代詞	人稱代詞, 指示代詞, 疑問代詞
8	連詞	
9	介詞	
10	助詞	結構助詞, 時態助詞, 時制助詞, 比況助詞, 表數助詞, 列舉助詞, 語氣助詞, 其他助詞
11	短語	量詞短語, 方位短語, 介詞短語, “的”字短語
12	主語	
13	賓語	雙賓語
14	補語	結果補語, 趨向補語, 可能補語, 程度補語, 情態補語, 數量補語, 介詞, 短語補語
15	疑問句	是非問句, 特指問句, 選擇問句, 正反問句
16	句式	主謂謂語句, “把”字句, “被”字句, 連動句, 強調句, 兼語句, 使役句, 存現句, 比較句, “連”字句
17	複句	並列複句: 承接複句, 遞進複句, 選擇複句, 注解複句
		偏正複句: 因果複句, 條件複句, 轉折複句, 讓步複句, 目的複句

『新編現代漢語』(張斌、齊滄揚等, 2002: 273-467)

Learners' Error Corpora of Chinese Searching Platform

About the site

This site is to search the error corpora of Chinese. The error corpora are collected by the team of K.Mochizuki, H.Sano, Y.M.Shen, S.Fukuda at the Tokyo University of Foreign Studies in Japan and collaborated with C.H.Wu at the National Taiwan Normal University in Taiwan

Contributors [[Click Here](#)]

Search with Keyword

Error: --> Revision:

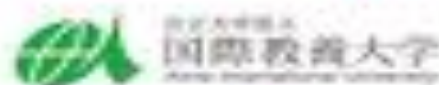
ex. 对、略、在、的

Search with Type

Search with category

The result will be shown here.

TUFS/NTNU/SISU/PEKING/LEEDS/AIU
 Learners Corpora of English,
 Chinese and Japanese



東京外國語大學
 科學研究費 基礎研究 (B)
 《英日漢語詞語基礎語彙的建構
 及以母語為基礎的英日漢語教學法的開發》
 (25284001 藍月金子代表 2013-2016年) *

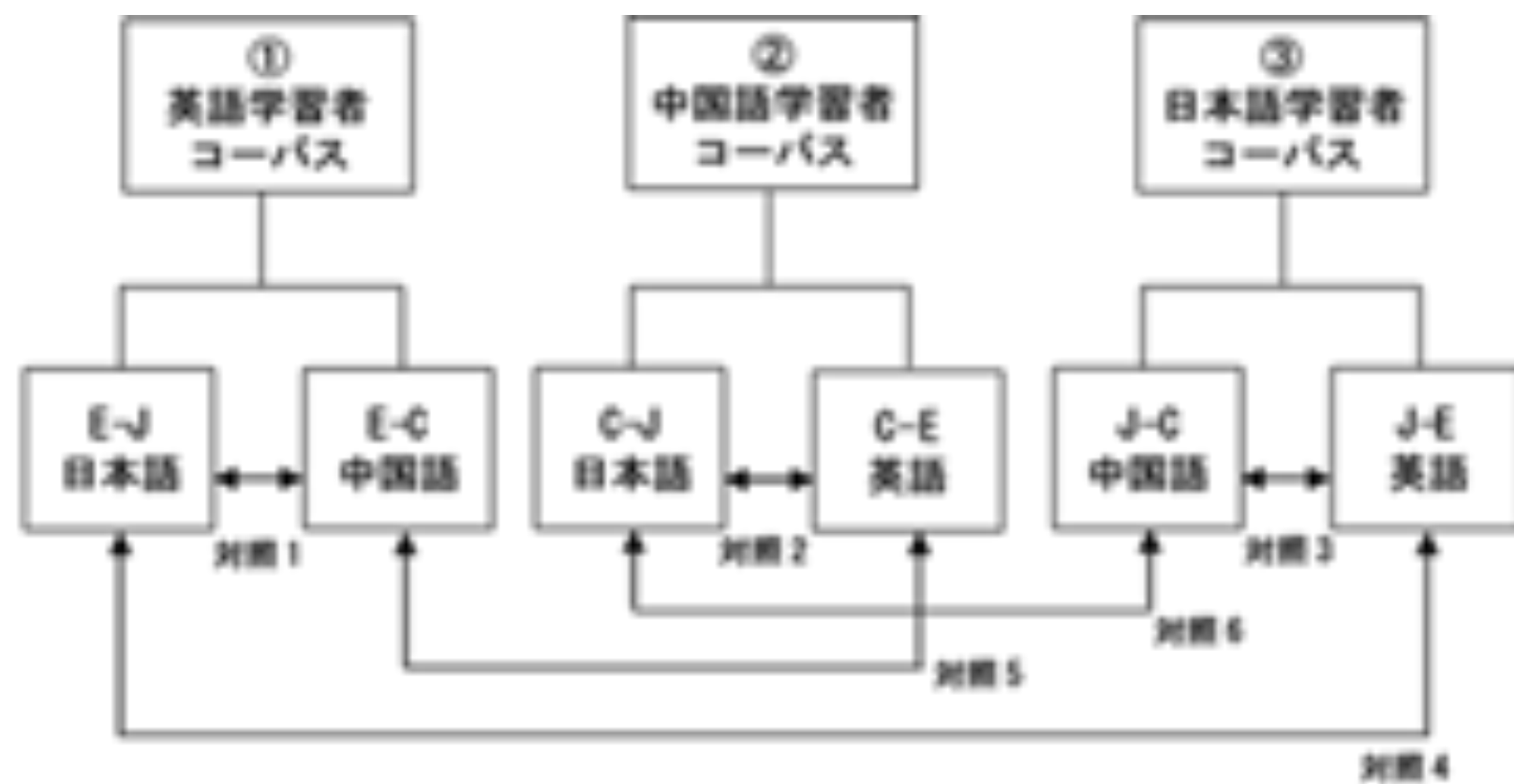
國立台灣師範大學邁向頂尖大學計畫*



(Ver.1.0)*

1) 学習言語

2) 学習者の母語



Case Study 1

**The acquisition of
Yi (—) + classifier**

- Japanese native speakers: **underuse of Yi + CL**
- English native speakers: **overuse of Yi + CL**

B1 and B2 level	Number of Chinese characters	Occurrence of YI + CL
TUFS (Japanese L1)	134,094	1490
NTNU (English L1)	151,823	385

$X^2(1)=150.03; p<.001$

B1 and B2 level	Underuse	Overuse	Others
TUFS (Japanese L1)	176 (98.32%)	2 (1.12%)	1 (0.56%)
NTNU (English L1)	12 (17.91%)	52 (77.67%)	3 (4.48%)

Error Examples (Japanese L1 Learners)

- (1) 生孩子是一件不简单的事。
- (2) 京都是一个很美丽的地方。
- (3) 东大和有一个很大的公园，附近也有一条小河。
- (4) 最近他在车站附近开了一家中餐馆。
- (5) 他从抽屉里拿出一本书，读了起来。
- (6) 我送给妈妈一罐咖啡豆，因为妈妈喜欢喝咖啡。

Error Examples (English L1 Learners)

- (7) 有什麼問題就跟我打 **一通** 電話吧。
- (8) 我們游完泳我計劃我們去電影院看 **一部** 電影。
- (9) 你開 **一個** 慶祝會的時候，我不能參加.....
- (10) 我在台北沒有發生 **一個** 大問題。
- (11) 然后我們可以在这个电视上看 **一个** 节目。

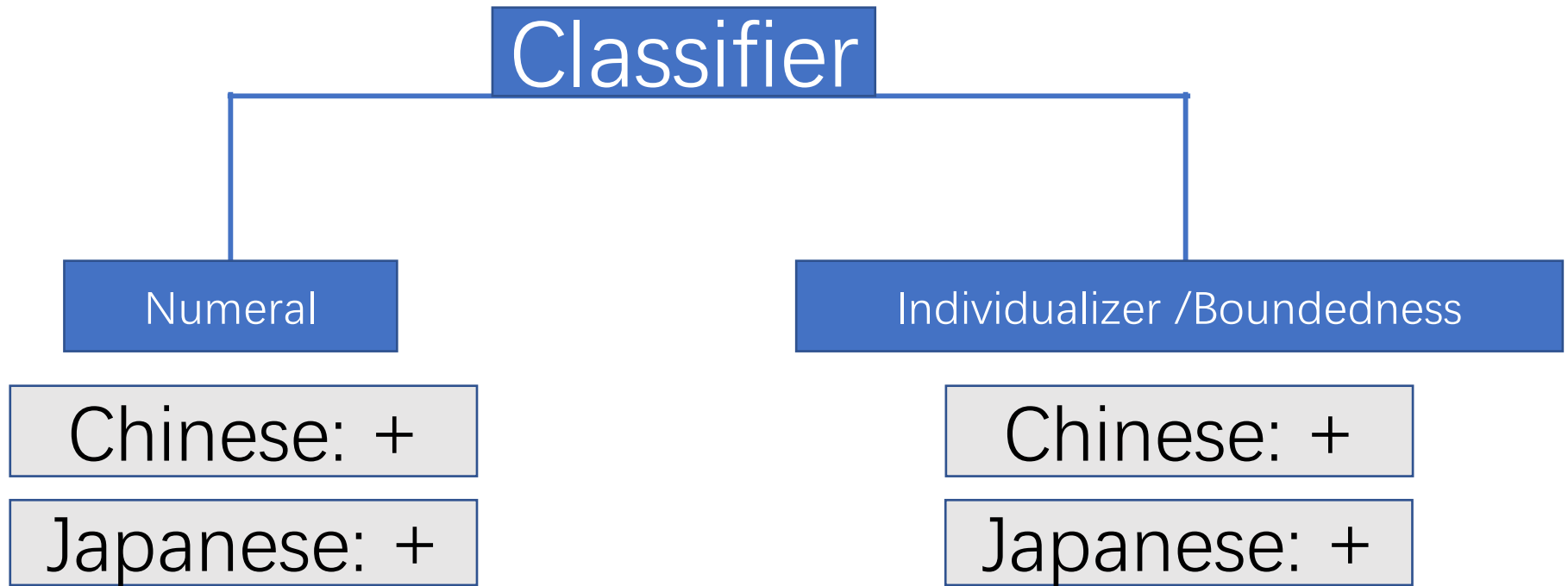
Discussion

- Bare nouns like cotton (棉花) , glass (玻璃) , book (书)
→ mass noun
- Classifier functions as a(n)
 - Individualizer (Okouchi 1985)
 - Unitizer (Lucy 1992)
- (12) a. *两书
b. 两本书
- (13) a. ??打破玻璃
b. 打破两块玻璃

Discussion

- Shen(沈家煊) (1995) discusses the interaction between “ Yi + Classifier” and the concept of ‘bounded’ and ‘unbounded’ events. Shen (1995) indicates that a “ Yi + Classifier” is necessary before a ‘bounded’ Noun Phrase(NP) in ‘Telic’ events.

Discussion



Case Study 2

**The acquisition of
completive aspect**

**-focus on
'verb- resultative complement'**

- Japanese native speakers: **underuse V-R**
- English native speakers: **overuse of V-R**

Frequency of Occurrence per 100,000 words

	Jp L1 Learners	Eng L1 Learners	Chinese Native Speakers ^[1]
a. V-到	393.98	523.93	3338.78
b. V-成	58.64	53.26	260.35
c. V-完	18.32	82.99	15.60

^[1] The Modern Chinese Corpus (现代汉语语料库检索) Institute of Applied Linguistics, Ministry of Education, China.
<http://www.cncorpus.org/index.aspx>

Error Examples (Japanese L1 Learners)

- (12) 如果你住在城市的话，你会比较容易**找→找到**工作。(2013_068_TUFS_CH_077)
(都市に住めば、仕事がすぐ見つかる。)
- (13) 如果买杯面的话，**吃→吃完**后我们可以扔掉容器。(2013_130_TUFS_CH_086)
(カップ麺を買えば、食べ終わったら容器を捨てればいい。)
- (14) 你们能体验把鲜鱼**做→做成**鱼糕。(2014_118_TUFS_CH_053)
(皆さんは、鮮魚をかまぼこにする経験をすることができます。)

Discussion

张黎 (2009)

王さんはコップを壊した。 (Result(Action))

小王打碎了杯子。 (Action+Result)

田中さんは卵を割れた。 (Result(Action))

田中捏碎了鸡蛋。 (Action+Result)

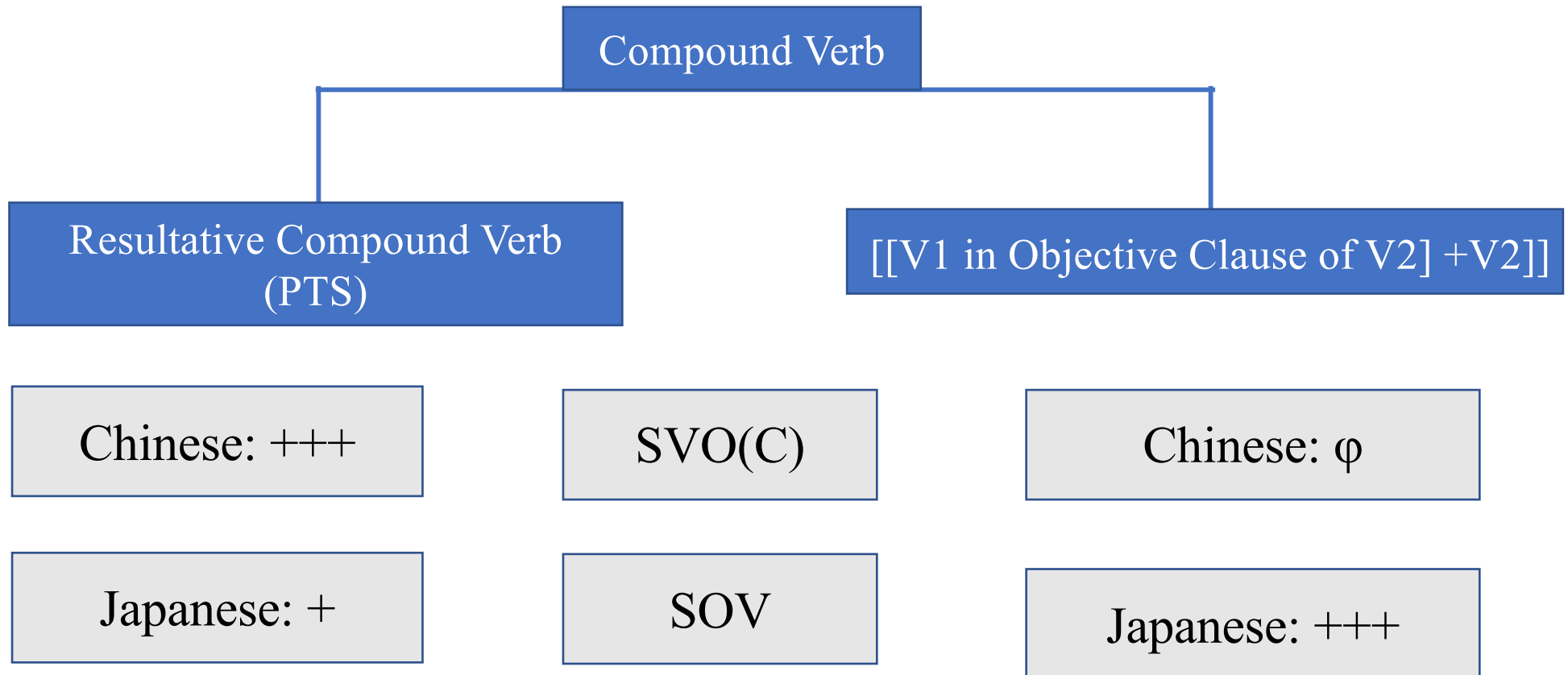
cf. Tai, J. (1984, 1985): The principle of temporal sequence

Japanese Aspectual Compound Verbs



Different Ways in which a Situation Unfolds

Discussion



Summary

- Native language influences the kinds of errors learners make
 - Incorrect analogies can lead to overuse of TL forms
 - Lack of salience in NL can lead to underuse of TL forms
- Language typology can help us understand error trends
- It is indispensable to explore the pedagogy of Chinese based on learners' native language to develop more efficient and advanced learning science

Thank you for your attention