

# Corpora of linguistic acceptability (CoLA) as a tool for objective determination of the difficulty of grammar points



Hiroki Nomoto & Rin Kuroda  
Tokyo University of Foreign Studies (TUFS)

1 December 2022    CLaSIC 2022@NUS

# Background: Grammatical knowledge in language teaching

- Grammatical knowledge constitutes an important aspect of proficiency.
- Teaching material development

Authors carefully plan the order of presenting grammar points, so that basic points are introduced before advanced ones.

- Assessment

Whether a learner has mastered certain grammar points is used as an indicator of his/her general proficiency level.

# Problem & Proposal

- Problem

Despite our frequent reference to it, the difficulty of different grammar points tends to be determined subjectively by individual teachers, but **not based on the results of objective research**.

- Proposal

We propose to make use of **corpora of linguistic acceptability (CoLA)** to examine the difficulty of grammar points.

# Outline

- Previous work on grammatical knowledge assessment (especially in the context of extensive reading)
- CoLA (corpus of linguistic acceptability)
- How to use CoLA to determine difficulty levels of grammar points
- MALINDO CoLA and its use in the teaching of Malay/Indonesian

# Previous work on the use of CoLA in language teaching

None

(to the best of our knowledge)

# Grammatical knowledge assessment in extensive reading

In grammar tests for measuring the effectiveness of extensive reading, existing tests are modified or used as they are.

1. Rodrigo et al. (2004)
  - a. University students learning Spanish
  - b. A grammar test consisting of 30 **multiple choice questions** from a test published by the Ministry of Education and Science of Spain.
2. Imamura (2008)
  - a. Japanese high school students learning English
  - b. A grammar test consisting of 35 **multiple choice questions** from STEP Eiken test: Grade 3 (15), Grade pre-2 (10) and Grade 2 (10)

# EPER Test

- The effectiveness of extensive reading is often assessed by means of **cloze tests**.
- The EPER Test developed by the Edinburgh Project on Extensive Reading is one such test.

## Sample test from <https://enpedia.rxy.jp/wiki/EPERテスト>

Hello, I am Yurupedia. I have had an account in Enpedia (1) 2017. I like Enpedia very much. Do you know (2) Enpedia was founded? I'm going to talk about history of Enpedia. The first Enpedia was founded in 2009. In 2010, the second Enpedia was founded. On January 1, 2013, the third and current Enpedia was founded. (3) the year finished, the logo of Enpedia was decided. A leaf of fatsia was (4) on it. In 2014, a problem happened. On February 2014, there were thousands of categories (5) there were only several hundred articles. At that time, (6) new categories with few articles was banned, but categories that (7) then were not deleted. About (8) months later, on May of the year the system was maintained. Then, because there were too many categories, the database of Enpedia was broken. So, categories that had four or (9) articles were deleted. Now, there are more than 30000 articles in Enpedia. Enpedia has not so many rules, so you will be (10) to enjoy in Enpedia. Why don't you join Enpedia?

# Assessment tests in extensive reading

- Examine general proficiency level
  - Difficult to separate grammatical knowledge from other abilities
- Adopt the level settings from existing tests
  - How are the levels determined in the first place?



# How are the levels determined?

Baba & Oikawa (1992)

	Easy	↔	Difficult
<b>Regularity:</b>	Regular items <i>searched</i>		Irregular items <i>sought</i>
<b>Simplicity:</b>	Simple items <i>the man</i>		Complex items <i>the tall man I met yesterday</i>
<b>Similarity:</b>	Similar to known items		Different from known items

Were these criteria discovered empirically? — Probably not...

# Proposal

We propose to make use of **corpora of linguistic acceptability (CoLA)** to examine the difficulty of grammar points objectively.

# What is CoLA?

- A corpus consisting of sentences illustrating particular grammar points together with their acceptability judgements.
- It was originally created in the field of natural language processing to evaluate to what degree computers have successfully learned the grammar of the target language without explicit instructions.

# Language processing by computers in the past

## Explicit instructions

### Grammatical rules

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$PP \rightarrow P NP \dots$

### Lexical items

Orthography: put

Phonology: /put/

Part of speech: V

Arguments: NP1 <NP2, PP>

Semantics: Agent <Theme, Location>



## Practical tasks

Text-to-speech conversion

Machine translation

Question answering

Text summarisation

Dialogue system



Linguistic knowledge acquired by computers

# Language processing by computers today

Massive input



Proposals of various evaluation metrics to study these issues



Practical tasks

Text-to-speech conversion

Machine translation

Question answering

Text summarisation

Dialogue system

What have computers learned?

Do they have the same kind and amount of linguistic knowledge as humans do?

# English CoLA (Warstadt et al. 2019)

- 10,657 sentences from 23 linguistics publications
- Examine how accurately computers can replicate the judgements

clc95	0	*	In which way is Sandy very anxious to see if the students will be able to solve the homework problem?
c-05	1		The book was written by John.
c-05	0	*	Books were sent to each other by the students.
swb04	1		She voted for herself.
swb04	1		I saw that gas can explode.

source

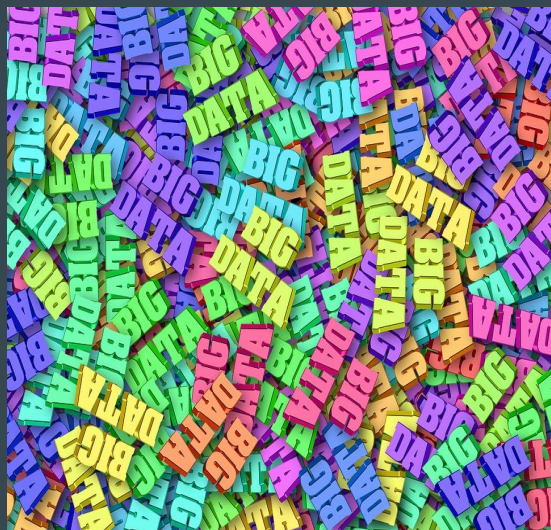
acceptability  
judgement (1 or 0)

original  
judgement

sentence

# CoLA can also be used to evaluate grammar learning by humans

Massive input



Practical tasks



Reading aloud foreign texts



Translation



Reading comprehension



Text summarisation



Conversation

What have human learners learned?

Do they have the same kind and amount of linguistic knowledge as native speakers do?

# Objective determination of difficulty levels of grammar points

## The basic idea

1. Conduct an **acceptability judgement test** with a group of learners using sentences from CoLA
2. Calculate the average score for each sentence\*



The higher the average test score is, the easier the relevant sentence is.

\*To determine the difficulty of a grammar point, replace “sentence” by “group of sentences illustrating the grammar point”.



# Example (N = 100)

Sentence (grammar point category)	Judgement	# response "1"	# response "0"	Score
A	1	100	0	100
B	0	10	90	90
C	1	50	50	50
D	0	90	10	10

Easy

Difficult

Data from CoLA

Acceptability judgement test results



# Subject groups

- The difficulty of a grammar point can vary depending on the **learner's L1**.  
e.g. “Advanced” for English speaker but “Intermediate” for Japanese speakers
- A crosslinguistic comparison based on the same CoLA will reveal such differences.  
e.g. Scores for sentence (/grammar point) A
  - 55 (L1 = English)
  - 60 (L1 = Mandarin)
  - 70 (L1 = Japanese)
  - 71 (L1 = Korean)

# Currently available CoLAs

Language	Dataset	Size
English	CoLA (Warstadt et al. 2019)	10,657 sentences
English	BLiMP (Warstadt et al. 2020)	67,000 minimal pairs
Italian	ItaCoLA (Trotta et al. 2021)	9,722 sentences
Mandarin Chinese	CLiMP (Xiang et al. 2021)	1,600 minimal pairs
Japanese	JCoLA (Someya & Oseki 2022)	369 minimal pairs; 2,323 sentences

# MALINDO CoLA Project (grant application under review)

- MALINDO = Malay + Indonesian  

## Data collection (data = sentences + acceptability judgements)

- Linguistics articles
- Translations of CoLA sentences in other languages + judgements by native speakers
- Constructed sentences by linguists + judgements by native speakers

## Data verification

- Conduct an acceptability judgement test with native Malay/Indonesian speakers (15 speakers each)
- Revise the judgements accordingly and publish the corpus

# Determination of difficulty levels using MALINDO CoLA

- Conduct an **acceptability judgement test** with several groups of learners using sentences from MALINDO CoLA
  - University students learning Malay/Indonesian in Japan (TUFS, KUIS, Nanzan, Ritsumeikan APU)
  - University students learning Malay/Indonesian in Singapore (NUS) (+ Malaysia, Indonesia)
  - Students at different course levels
- Calculate the average score for each sentence/grammar point



The higher the average test core is, the easier the relevant sentence/grammar point is.

- The proposed study will also reveal
  - Grammar points whose acquisition is affected by L1 and those invariant regardless of L1
  - The order of acquisition

# Conclusions

- The difficulty of a grammar point can be determined objectively by using CoLA.
- The difficulty levels thus determined can be used in
  - teaching material development (textbooks, graded readers, etc.)
  - grammatical knowledge assessment
  - and perhaps more

**Collaborators** for MALINDO CoLA development are welcome!  
(provided that the grant application passes)

# References

- Baba, T. & K. Oikawa. 1992. Gakushubunpou nitsuite naniga iware, naniga wakatteiruka [What has been said and what is known about learner's grammar]. In K. Kanatani (ed.) *Gakushubunpou: Bunpousho, bunpoukyouiku no hataraki o saguru*, 57–112. Tokyo: Kagensha.
- Imamura, K. 2008. The effects of extensive reading for Japanese high school students on their reading and listening abilities, vocabulary and grammar. *ARELE: Annual Review of English Language Education in Japan* 19: 11–20.
- Rodrigo, V., S. Krashen, & B. Gribbons. 2004. The effectiveness of two comprehensible-input approaches to foreign language instruction at the intermediate level. *System* 32(1): 53–60.
- Someya, T. & Y. Oseki. 2022. Nihongoban CoLA no kouchiku [Building a Japanese CoLA]. In *Proceedings of the Twenty-Eighth Annual Meeting of the Association for Natural Language Processing, 1872–1877*.
- Trotta, D., R. Guarasci, E. Leonardelli & S. Tonelli. 2021. Monolingual and cross-lingual acceptability judgments with the Italian CoLA corpus. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2929–2940. Association for Computational Linguistics.
- Warstadt, A., A. Singh & S. R. Bowman. 2019. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics* 7: 625–641.
- Warstadt, A., A. Parrish, H. Liu, A. Mohananey, W. Peng, S. Wang & S. R. Bowman. 2020. BLiMP: The benchmark of linguistic minimal pairs for English. *Transactions of the Association for Computational Linguistics* 8: 377–392.
- Xiang, B., C. Yang, Y. Li, A. Warstadt & K. Kann. 2021. CLiMP: A benchmark for Chinese language model evaluation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 2784–2790. Association for Computational Linguistics.