# 10 Can you beat TOROBO-kun?
## ~ Evaluating the naturalness of sentences using language models ~

## Abstract

We are participating in the "Can a robot get into the University of Tokyo?" project of the National Institute of Informatics, focusing on how to tackle English exams. Succeeding on English exams requires not only English knowledge but also "common sense knowledge" that humans take for granted but computers do not possess. For example, there are problems that require estimation of human emotion and the causal relationships of events. With our common sense reasoning techniques developed through research on dialogue and machine translation systems, we have for the first time achieved a higher score than the average human score on a Mock Center Test.

## Objective: Auto-answering on English exams

• "Common sense knowledge" required to solve English exam questions

• Key point: How to model such common sense?

---

Parker: I hear that your father is in hospital.
Brown: Yes, and he has to have an operation next week.
Parker: [ 19 ] Let me know if I can do anything.
Brown: Thanks a lot.

① Exactly, yes.
② No problem.
③ That's a relief.
④ That's too bad.

> It is most difficult for computers to answer such questions that are not appeared in textbook!

(Source: Yozemi Mock Center Test of the Educational Foundation Takamiya Gakuen)
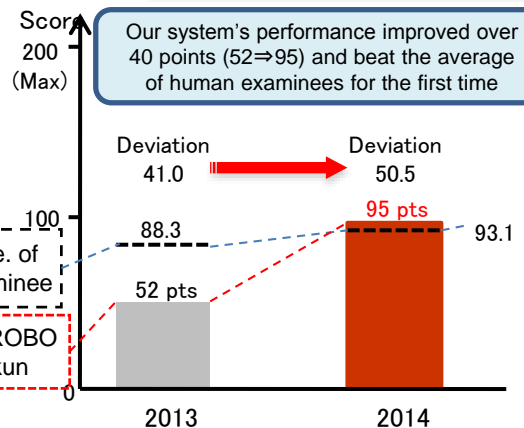
---

Our system leverages techniques of dialogue processing and machine translation to answer test questions

E.g.
Evaluating naturalness of dialog flow

Points to consider:
・ Dialogue act (offer, thanks…)
・ Sentiment polarity (pos./neg.)

---

| 1 | Pronunciation | 1A | Pronunciation | Short-sentence questions (words, a few sentences) Points: 99 |
| | | 1B | Accent | |
| 2 | Grammar Dialogue Ordering | 2A | Grammar/Vocabulary | |
| | | 2B | Dialogue completion | |
| | | 2C | Word ordering | |
| 3 | Reading (context understanding) | 3A | Word meaning estimation | |
| | | 3B | Irrelevant sentence finding | |
| | | 3C | Opinion summarization | |
| 4 | Reading (info. processing) | 4A | Graph/Pictures/Tables | |
| | | 4B | Reading practical documents | |
| 5 | Reading (info. understanding) | | | Long-sentence questions Points: 101 |
| 6 | Reading(article) | | | |

Our system's performance improved over 40 points (52⇒95) and beat the average of human examinees for the first time

Score
200 (Max)

Deviation 41.0 ⇒ Deviation 50.5

Ave. of examinee — 88.3 ⇒ 93.1

TOROBO -kun — 52 pts ⇒ 95 pts

100

0

2013    2014

## Related works

[1] Ryuichiro Higashinaka, Hiroaki Sugiyama, Hideki Isozaki, Genichiro Kikui, Kohji Dohsaka, Hirotoshi Taira, Yasuhiro Minami, "Taking the English exam in the National Center Test for University Admissions", in Proc. t he 21st Annual Meeting of The Association for Natural Language Processing (NLP2015), 2015. (in Japanese)

[2] Ryuichiro Higashinaka, Hiroaki Sugiyama, Hideki Isozaki, Genichiro Kikui, Kohji Dohsaka, Hirotoshi Taira, Yasuhiro Minami, "Taking the English exam for the "Can a robot get into the University of Tokyo?" project?", NTT Technical Review, 2015. (to appear)

## Contact

**Hiroaki Sugiyama** Interaction Research Group, Innovative Information Laboratory
E-mail : sugiyama.hiroaki(at)lab.ntt.co.jp