

The Winograd Schema Challenge

By Ernest Davis, Leora Morgenstern, and Charles Ortiz

Winograd Schemas

A Winograd schema is a pair of sentences that differ in only one or two words and that contain an ambiguity that is resolved in opposite ways in the two sentences and requires the use of world knowledge and reasoning for its resolution. The schema takes its name from a well-known example by Terry Winograd

The city councilmen refused the demonstrators a permit because they [feared/advocated] violence.

If the word is "feared", then "they" presumably refers to the city council; if it is "advocated" then "they" presumably refers to the demonstrators.

In his paper, "The Winograd Schema Challenge" Hector Levesque (2011) proposes to assemble a set of such Winograd schemas that are

- easily disambiguated by the human reader (ideally, so easily that the reader does not even notice that there is an ambiguity);
- not solvable by simple techniques such as selectional restrictions;
- Google-proof; that is, there is no obvious statistical test over text corpora that will reliably disambiguate these correctly.

The set would then be presented as a challenge for AI programs, along the lines of the Turing test. The strengths of the challenge are that it is clear-cut, in that the answer to each schema is a binary choice; vivid, in that it is obvious to non-experts that a program that fails to get the right answers clearly has serious gaps in its understanding; and difficult, in that it is far beyond the current state of the art.

A contest, entitled [the Winograd Schema Challenge](#) was run once, in 2016. At that time, there was a cash prize offered for achieving human-level performance in the contest. Since then, the sponsor has withdrawn; therefore **NO CASH PRIZES CAN BE OFFERED OR WILL BE AWARDED FOR ANY KIND OF PERFORMANCE OR ACHIEVEMENT ON THIS CHALLENGE.**

Pronoun Disambiguation Problems

As discussed in (Morgenstern, Davis, and Ortiz, 2016), the first round of the 2016 offering of the Winograd Schema Challenge was a collection of Pronoun Disambiguation Problems (PDPs). These are texts, found in natural literary sources, containing an pronoun that requires the use of commonsense knowledge to disambiguate.

Collection of Winograd Schemas

A collection of 150 Winograd schemas has been developed and published.

- [Collection in XML](#)
- [Collection in HTML](#). This is the original published form of the collection. It includes some discussion at the beginning, and some material that does not fit well in the more rigid XML file.

Both versions of the collections are licenced under a [Creative Commons Attribution 4.0 International License](#).

For comparison, there is also a small [collection](#) of examples that, though interesting, in one way or another fail to meet the bar of WS schemas. To avoid confusion, these are placed on a separate page.

Winograd Schemas in other Languages

[Translation of 12 WSs into Chinese](#) (translated by Wei Xu).

Translations into Japanese, by Soichiro Tanaka, Rafal Rzepka, and Shiho Katajima

[Translation changing English names to Japanese PDF](#) [HTML](#)

[Translation preserving English names PDF](#) [HTML](#)

[Translation into French](#), by Pascal Amsili and Olga Seminck

[Winograd Schemas in Portuguese](#) by Gabriela Melo, Vinicius Imaizumi, and Fábio Cozman.

[Mandarinograd: A Chinese Collection of Winograd Schemas](#) by Timothée Bernard and Ting Han, LREC-2020.

Translation into Hebrew, by Vered Shwartz. [HTML](#) [JSON](#).

[Translation into Hungarian](#) by Noémi Vadász and Noémi Ligeti-Nagy.

[Translation into Russian](#) by Tatiana Shavrina et al.

The 2016 running of the Winograd Schema Challenge

Summary: The Winograd Schema Challenge was run on July 11, 2016 at IJCAI-16. There were four contestants. The first round of the challenge was a collection of 60 PDP's. The highest score achieved was 58% correct, by Quan Liu, from University of Science and Technology, China.. Hence, by the [rules of that challenge](#), no prizes were awarded, and the challenge did not proceed to the second round.

- [Short report on the event](#). *AI Magazine*, Fall 2017, pp. 97-98.
- [Sample input](#) (available to contestants before the event).
- [Sample output](#) (corresponding to sample input)
- [Challenge problems for WSC 2016](#)
- [Human subject tests \(Draft\)](#)
- [Commonsense Knowledge Enhanced Embeddings for Solving Pronoun Disambiguation Problems in Winograd Schema Challenge](#) Quan Liu et al., 2016.

More recent results

A comprehensive survey of the state of the art as of April 2020 is given in [A Review of Winograd Schema Challenge Datasets and Approaches](#) by Vid Kocijan, Thomas Lukasiewicz, Ernest Davis, Gary Marcus, and Leora Morgenstern. arXiv 2004.13831. April 2020.

Publications

David Bender, [Establishing a Human Baseline for the Winograd Schema Challenge](#). *MAICS* 2015

Ernest Davis, [Winograd Schemas and Machine Translation](#), arXiv:1608.01884, August 2016.

Hector Levesque, [The Winograd Schema Challenge](#), *Commonsense-2011*.

Hector Levesque, Ernest Davis, and Leora Morgenstern, [The Winograd Schema Challenge](#), KR-2012. An expanded version of the previous item.

Hector Levesque, [On Our Best Behaviour](#), IJCAI Research Excellence Award Presentation, 2013.

Leora Morgenstern, E. Davis, and Charles Ortiz, [Planning, Executing, and Evaluating the Winograd Schema Challenge](#), *AI Magazine* Spring 2016. [Article in the magazine](#)

Acknowledgements

Thanks to Hector Levesque, Gary Marcus, Ray Jackendoff, David Bender, Wei Xu, So Tanaka, and Rafal Rzepka.