# MaTaCOp

## The Map Task Corpus of The Open University of Israel

# Background

MaTaCOp is a large scale project which strives to compile a Map Task corpus in Hebrew, which consists of spoken task-oriented dialogues following the Human Communication Research Centre (HCRC) standards (Anderson et al., 1991).

MaTaCOp is destined for a wide range of (spontaneous) speech and behavioral sciences, as well as to speech technologies purposes, and is released for research purposes only.
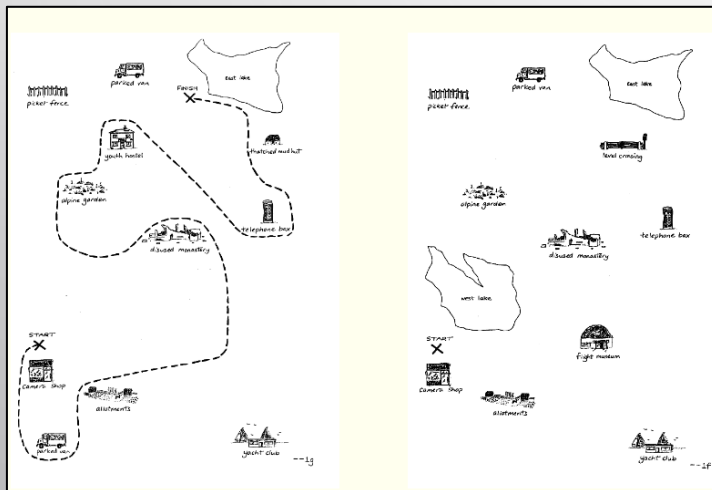
# Motivation

Every language in which a rich linguistic corpus like Map Task was created led to a wealth of studies based on the corpus.

Other groups around the world compiled similar corpora in other languages. For example:

- German: HaMaTaC, BeMaTaC
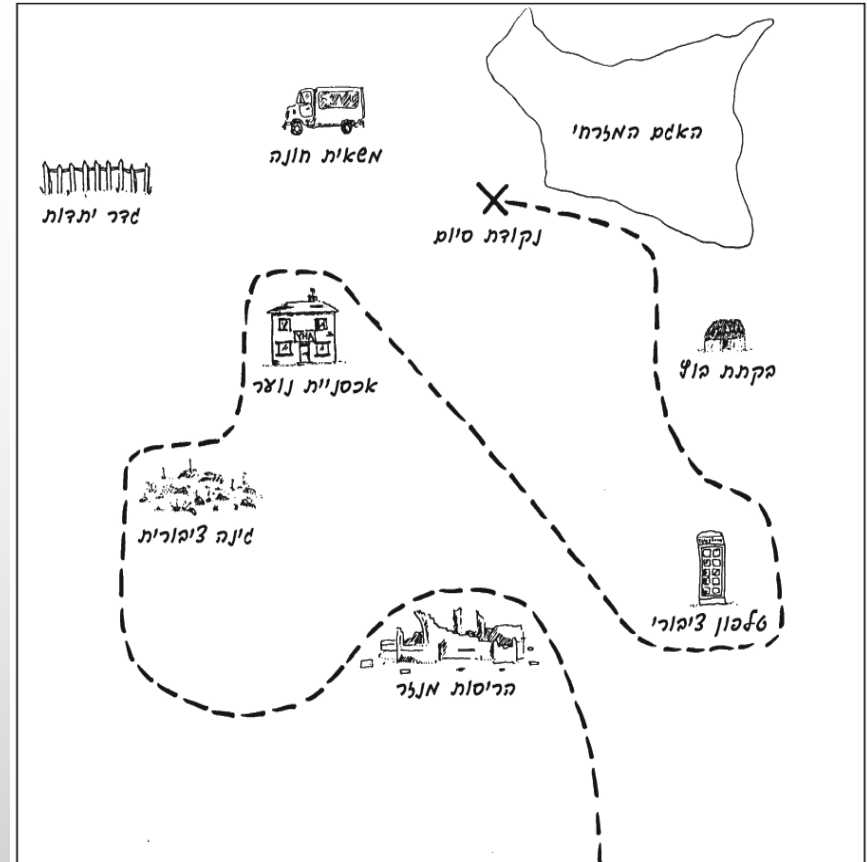- Portuguese: CORAL
- French: MAPTASK-AIX

# Design

The corpus uses the map task design in which speakers must collaborate verbally to reproduce on one participant's map a route printed on the other's.



Screenshot taken from http://groups.inf.ed.ac.uk/maptask/

# Design

We used two pairs of maps from the original set (Anderson et al., 1991), with translation of the milestones into Hebrew .

# Recordings and Participants

- The Hebrew map task corpus consists of 32 speakers who participated in two sessions, once as a "leader" and once as a "follower", thus creating a set of 32 dialogues.

- The recordings were taken during September 2015-January 2016.

- All speakers are fluent in Hebrew and use Hebrew as their main communication language since childhood, but not necessarily native speakers.

# MaTaCOp
## The Hebrew Map Task Corpus
## of the Open University of Israel

To the best of our knowledge there is no Map Task designed corpus in Hebrew. Thus, this project is a pioneer map task corpus in Hebrew.

We hope it will serve as a platform for rich research on Hebrew speech and language.

# Recording setup

All recording were taken according to the same setup.

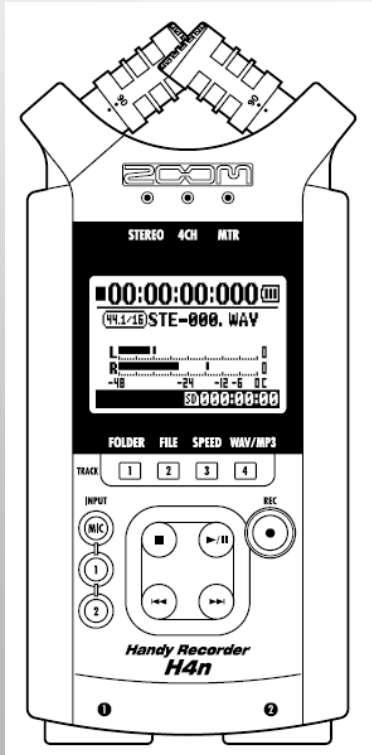The following parameters were strictly kept:

- Distance between participants.

- No air-conditioning.

- Closed windows and door.

- Computers were shut off.

- A carpet under the participants' chairs.

- Microphone close to the speaker's mouth but not touching it.

- Comfortable adjustment of the headset ("Madonna" type) to the speaker.

# Recording device

The recording device was H4N (ZOOM-NA.COM).

- Activation via batteries.

- Two paths stereo, with two passive microphones, one per speaker.

- 96kHz sampling rate.

- 24 bit.

- WAV Format.

- No signal processing.

# Recording Equipement

# Demo

# CREDIT

The corpus was compiled at The [Department of Mathematics and Computer Science](#) and the [Research Center for Innovation in Learning Technologies](#) at the [Open University of Israel](#).

Please cite the following source in any published work which is based on the corpus:

Azogui, J., Lerner, A., and Silber-Varod, V., The open university of Israel Map Task Corpus (MaTaCOp). Available at: www.Openu.Ac.Il/matacop.

# References

- Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., Mcallister, J., Miller, J., Sotillo, C., Thompson, H. S. and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, *34*, pp. 351-366.