Plan Overview

A Data Management Plan created using DMPonline

Title: The influence of typology and microvariation in L3: The case of the interplay of Cantonese, Mandarin and English in comparative constructions

Creator: Chun Yan Tung

Principal Investigator: Chun Yan Tung

Data Manager: Chun Yan Tung

Project Administrator: Chun Yan Tung

Affiliation: The University of Sheffield

Funder: Arts and Humanities Research Council (AHRC)

Template: AHRC Data Management Plan

Project abstract:
This study examines language combinations that have not been widely studied within L3 literature: Cantonese, Mandarin and English. Specifically, the comparative structures in the three languages will be studied, offering a unique platform to test two underpinning questions on acquisition studies: the learnability problem by Poverty of the Stimulus and the nature of transfer. These two questions are hitherto unexplored in existing L2 and L3 models; as such results collected could shed light on what actually transfers in the context of second and third language acquisition.

ID: 157367

Start date: 01-10-2023

End date: 01-10-2026

Last modified: 11-09-2024

Grant number / URL: https://wrocah.ac.uk/funding/current-funding-for-2019-2023-students/

The influence of typology and microvariation in L3: The case of the interplay of Cantonese, Mandarin and English in comparative constructions

Data Summary

1. Briefly introduce the types of data the research will create. Why did you decide to use these data types?

Data type	File extension	Source: collection methods & timeframe	Digital data size (approx.)	Conditions of use
1. Test tokens for all the grammaticality judgement tasks to be administered in Year 2.	.txt	Tokens to be designed between September 2024 and December 2024. Tokens will be refined after a pilot test is conducted in or before November 2024.	20 MB	The test tokens will not be shared before the grammaticality judgement tasks are administered. The test tokens will only become publicly available once a paper using the respective test tokens is published. Publishing the test tokens does not and will not lead to any copyright issues.
2. Grammaticality judgement task data held in spreadsheets	.csv	Approximately 100 participants online tests through Gorilla -- will be collected approx. from May 2025 to July 2025	5 MB	Any identifiable data will not be shared. Each participant will be assigned a unique number ID, which will be used to link their responses throughout the project. These ID and responses will be shared with other researchers. Data can be accessed with ORDA.
3. Analysis of the test results o f the grammaticality judgement task spreadsheets	.csv	Approximately 100 participants online tests through Gorilla analysis will be conducted in August 2025.	1 MB	Any identifiable data will not be shared. Each participant will be assigned a unique number ID, which will be used to link their responses throughout the project. These IDs and responses will be shared with other researchers. The analysis of the data will be shared.

Data Collection

2. Give details on the proposed methodologies that will be used to create the data. Advise how the project team selected will be suitable for the data/digital aspects of the work, including details of how the institution's data support teams may need to support the project

1. The test tokens are to be created for at least three tests and will be stored separately on Word documents. The Word documents will be labelled according to the specific tests that they are used for.

2. Test results of grammaticality judgement tasks will be collected online using Gorilla as the platform. Data will be pseudonymised, and participants will be assigned a numerical ID. A copy of the data will be created and saved separately in a separate, encrypted folder on the university's X drive in addition to copies of the data on Google Drive and my own password-protect storage device.

Short-term Data Storage

3. How will the data be stored in the short term?

Most data will be stored on the University's Google Drive with two-factor authentication and the researcher's laptop with enhanced security (password). In addition, copies of the research data are also stored on the University's X Drive.
 The folder location plan is recorded on a README file as follows:
Thesis

- Grammaticality judgement task test 1_control_test tokens
- Data
- Grammaticality judgement task test 2_L2A_test tokens
- Data
- Grammaticality judgement task test 2_L3A_test tokens
- Data
- README

- Grammaticality judgement task_1_Group 1_L1
- Data
- Group 1_L1_Results
- README
- Analysis
- Group 1_L1_Analysis
- README

- Grammaticality judgement task_1_Group 2_L1
- Data
- Group 2_L1_Results
- README
- Analysis
- Group 2_L1_Analysis
- README

- Grammaticality judgement task_1_Group 1_L2
- Data
- Group 1_L2_Results
- README
- Analysis
- Group 1_L2_Analysis
- README

- Grammaticality judgement task_1_Group 2_L2
- Data
- Group 2_L2_Results
- README
- Analysis
- Group 2_L2_Analysis
- README

- Grammaticality judgement task_1_Group 1_L3
- Data
- Group 1_L3_Results
- README
- Analysis
- Group 1_L3_Analysis
- README

- Grammaticality judgement task_1_Group 2_L3
- Data
- Group 2_L3_Results
- README

- Analysis
- Group 2_L3_Analysis
- README

- Grammaticality judgement task_1_Group 1_L2A
- Data
- Group 1_L2A_Results
- README
- Analysis
- Group 1_L2A_Analysis
- README

- Grammaticality judgement task_1_Group 2_L2A
- Data
- Group 2_L2A_Results
- README
- Analysis
- Group 2_L2A_Analysis
- README

- Grammaticality judgement task_1_Group 1_L3A
- Data
- Group 1_L3A_Results
- README
- Analysis
- Group 1_L3A_Analysis
- README

- Grammaticality judgement task_1_Group 2_L3A
- Data
- Group 2_L3A_Results
- README
- Analysis
- Group 2_L3A_Analysis
- README

- Grammaticality judgement task_1_Group 1_L2A_morpheme
- Data
- Group 1_L2A_morpheme _Results
- README
- Analysis
- Group 1_L2A_morpheme _Analysis
- README

- Grammaticality judgement task_1_Group 2_L2A_morpheme
- Data
- Group 2_L2A_morpheme_Results
- README
- Analysis
- Group 2_L2A_morpheme_Analysis
- README

3a What backup will you have in the in-project period to ensure no data is lost?

The data will be regularly backed-up in the university's Google Drive and X drive (at least once every two days). The data is also kept in an external storage device for an offline backup.

Long-term Data Storage

4. How the data will be stored in the long term?

As per the conditions of my funding, I am obliged to make my data accessible for a minimum of 10 years after the completion of my PhD. In order to comply with the UKRI's recommendation, I will be storing my data in a repository. If the chosen repository requires it, I will make changes to the naming format so that it adheres to the guidelines.

I will store the methodology for data collection, processing, and analyses along with metadata. Any codes or programs used for the results and/or analysis will be made available. The grammaticality judgement tests data will be available but anonymised. I believe that my data could be helpful for future research and therefore, accessibility is crucial.

4a. Where have you decided to store it, why is this appropriate?

If a subject-specific repository is unavailable, the university suggests using ORDA, their own data repository. The guidelines require that the data be anonymous and non-sensitive, which my data will adhere to.

4b. How long will it be stored for and why?

My funding from AHRC/UKRI requires that my data be stored for a minimum of 10 years following the conclusion of my PhD.

4c. Costs of storage – why are these appropriate? Costs related to long term storage will be permitted providing these are fully justified and relate to the project Full justification must be provided in Justification of Resources (JoR)

Since the data is expected to be less than 250GB, there won't be any storage costs.

Data Sharing

5. How the data will be shared and the value it will have to others

Information regarding the methodologies will be stored in a repository for easy access by other researchers. This set of data is particularly valuable as the process of testing an under-researched grammatical property (i.e., the comparative construction) because this set of data would provide insights to future researchers as to how this property can be carefully teased apart for experiments with different variables controlled for. Also, this set of data is the first set of data in the field of Second and Third Language Acquisition for the testing for this particular grammatical property, which means that researchers could refer to the questions in the set of data to study other language pairings in the future.

5a. How the data will enhance the area and how it could be used in the future?

The data collected can enhance research in various ways. It can provide language teachers and practitioners with insights on understanding the directionality of transfer of acquired linguistic knowledge of grammatical properties from previously acquired languages. The data collected can shed light on the effects of Poverty of Stimulus and the (non-)facilitative effects of the first and/or second language.

5b. Releasing the data – advise when you will be releasing and justify if not releasing in line with AHRC guidelines of a minimum of three years. If the data will have value to different audiences, how these groups will be informed?

I plan to release the data within a year of my PhD thesis being published. It is common practice to have an embargo before the publication of research findings.

5c. Will the data need to be updated? Include future plans for updating if this is the case.

It is unlikely that the data will be updated. Additionally, the experimental design studies have a limited time frame, which means that there will be no new data to be collected after the allocated time.

5d. Will the data be open or will you charge for it? Justify if charging to access the data

Access to the data will be free through ORDA.

5e. Financial requirements of sharing – include full justification in the JoR

Not applicable.

Ethical and Legal Considerations

6a. Any legal and ethical considerations of collecting the data

To protect the privacy of participants, the data to be collected will be made anonymous, and Participants will be asked to provide consent for their participation, and they can revoke any information they have provided. We will also seek consent for storage, collection, usage, and methods. After the project, the personal details of the interviewees will be deleted. Data will be safeguarded through password protection, encryption, and secure servers.

6b. Legal and ethical considerations around releasing and storing the data – anonymity of any participants, following promises made to participants

The information gathered from surveys and interviews will have identifying details removed to protect the privacy of participants. Consent will be obtained from participants before storing and using their data. The depository will be available for other researchers to access the collected data.