

# Data Management Plan C14/21/038

## ADMIN DETAILS

**Project/Grant number:** C14/21/038

**Principal Investigator / Researcher:** Dorien Van De Mieroop

**Principal Investigator/Researcher ID:** 0000-0002-1704-2165

**Data contact:** Dorien Van De Mieroop (PI)/Melina De Dijn (PhD-student)

**Institution:** KU Leuven

## 1. GENERAL INFORMATION

### **Name of the project lead (PI)**

Dorien Van De Mieroop

### **C1-C2 Project number & title**

Project number: C14/21/038

Project title: A linguistic analysis of indexical markers of identity in Flemish job interviews in a changing recruitment context: A sociolinguistic-pragmatic mixed methods study.

### **Project description**

From the perspective of linguistic pragmatics and sociolinguistics, the construction of identity tends to be analyzed by, among others, drawing on the principle of indexicality which connects linguistic utterances to extra-linguistic reality. On the one hand, languages contain particular deictic elements (e.g. pronouns) which almost automatically entail some – locally highly variable – indexical force. On the other hand, any linguistic element may index particular context-related social meanings. A specific example concerning the Belgian Dutch linguascape is the fluid social meaning of different language varieties such as Colloquial Belgian Dutch and Belgian Standard Dutch. In this project, these indexical markers will be studied from a pragmatic and sociolinguistic perspective by means of a mixed methods approach. As data, Flemish job were chosen because these constitute an interesting context: i.e. they have changed a lot in the last decades in terms of the identities that recruiters and candidates construct, in particular because of the western workplace's shift to the New Work Order as well as the emergence of the War for Talent in some segments of the recruitment context.

## 2. DATA DESCRIPTION

### 2.1. Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing data

2.2. What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.

#### A. EXISTING CORPUS

*The existing data is a reference corpus containing 31 job interviews. They were collected by the PI who obtained SMEC-approval for this data (Ethical review (SMEC) approval numbers: G- 2015 10 371 and G- 2018 10 1354).*

Type of data	Format	Volume	How created?
Existing corpus of job interviews	Docx Audio files Video files	About 25 GB	We have two types of interviews which were created in different ways: <ol style="list-style-type: none"><li>1. interviews that were <u>audiotaped</u> using an audio-recorder;</li><li>2. interviews that were <u>videotaped</u> using a camera or smartphone.</li></ol> These interviews are fully transcribed <sup>1</sup> using conversation analytic transcription symbols (Jefferson 1984).

#### B. NEW DATA COLLECTED FOR THE PROJECT

*For the new data collection procedure consisting of the recording of job interviews followed by SRM-interviews, the PI and PhD-student will apply for SMEC-approval in May 2022. When the experiment is prepared, another SMEC-approval will be applied for. The rest of the data collection is still covered by the SMEC-approvals obtained earlier (including a few amendments) as described above.*

---

<sup>1</sup> The interviews were transcribed by students in the context of their master thesis research. The involvement of students in the data collection was inscribed in the SMEC-approval procedure.

Type of data	Format	Volume	How created?
New corpus of <i>War for Talent</i> vs <i>War for Jobs</i> -job interviews	Video files (mp4)	About 60GB (under construction)	<ul style="list-style-type: none"> <li>We have two types of video files which were created in different ways: <ol style="list-style-type: none"> <li>1. Job interviews that were videotaped using a <u>camera or smartphone</u>;</li> <li>2. Job interviews that were organized via Zoom, Skype, Teams, etc. and that were recorded through the <u>videoconferencing-software</u>.</li> </ol> </li> <li>Most of these job interviews are fully transcribed<sup>2</sup> using conversation analytic transcription symbols (Jefferson 1984). These transcriptions will be completed on a rolling basis (meaning, each time after new recordings are made).</li> </ul>
Coded transcripts of job interviews	.xlsx .r .tsv	< 1 GB	The job interviews will be coded for certain linguistic elements in an excel file. Analyses of these coded elements will be done in R.
Questionnaire	Docx .xlsx .r .tsv	< 1 GB (under construction)	<ul style="list-style-type: none"> <li>The recruiters taking part in this study filled out/will fill out a <u>questionnaire</u> with an assessment of the job interviews in terms of the War for Talent-context.</li> <li>Their answers will be structured in <u>an excel spreadsheet</u> to allow for analysis.</li> <li>Analyses will be done in R.</li> </ul>
Ethnographic interviews with recruiters	Audio files video files	About 3GB (under construction)	<p>We have two types of interviews which were/will be created in different ways:</p> <ol style="list-style-type: none"> <li>1. interviews that were audio- or videotaped using a <u>smartphone, audio-recorder or camera</u>;</li> <li>2. interviews that were organized via Zoom, Skype, Teams, etc. and that were recorded through the <u>videoconferencing-software</u>.</li> </ol>
SRM-interviews with recruiters	audio files video files	About 2GB (under construction)	<p>We have two types of SRM-interviews which were/will be created in different ways:</p> <ol style="list-style-type: none"> <li>1. interviews that were videotaped using a <u>camera or smartphone</u>;</li> <li>2. interviews that were organized via Zoom, Skype, Teams, etc. and that were recorded through the <u>videoconferencing-software</u>.</li> </ol>
Questionnaires collected during experiment	.xlsx .r .tsv	< 1 GB (to be collected)	During a large-scale experiment approx. 180 master students will fill out a questionnaire. Their answers will be collected in an excel file. Analyses will be done in R.

<sup>2</sup> The interviews were transcribed by students in the context of their master thesis research. The involvement of students in the data collection was inscribed in the SMEC-approval procedure.

### **3. ETHICAL AND LEGAL ISSUES**

**3.1. Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. Be aware that registering the fact that you process personal data is a legal obligation.**

Yes, all the project data contains personal information and the procedure for collecting and storing data was approved by SMEC. We would like to emphasize that we are extremely careful with storage of our data (e.g. OneDrive + KU Leuven authenticator).

**3.2. Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s).**

Yes there are ethical issues concerning this project. We have obtained **ethical approval** from SMEC (approval numbers: G- 2015 10 371 and G- 2018 10 1354) for the first part of the data collection and will apply for SMEC-approval for the new procedures (i.e. (1) the collection of job interviews + SRM interviews and (2) the experiment) after May 2022. Following SMEC-approval we collect informed consent forms from all participants for all interviews we record and if we use visual representations of participants in publications we use line drawings to protect their privacy. The use of these line drawings is also incorporated in the informed consent form that the participants sign, so they have explicitly agreed to the use of the visual data in this way. Since the start of the Covid-19 crisis, many job interviews take place via videoconferencing-software which makes collection of paper contracts less evident: a SMEC-amendment allows us to record the consent of the participants instead of working with paper contracts. We will describe our strategies to protect the privacy of the participants in our experiment in our new SMEC-approval applications (to be completed by May 2022 for the job interviews + SRM interviews, and in a later stage for the experiment).

**3.3. Does your research possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?**

NA

**3.4. Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions regarding reuse and sharing are in place?**

Yes, since this data is highly privacy sensitive. As stipulated in the informed consent forms that the participants sign, the data cannot be shared with anyone other than the researchers involved.

## 4. DOCUMENTATION AND METADATA

### 4.1. What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?

We will create a report describing the data used for this project that also makes clear how the different elements in the dataset belong together. A quick overview of the contents of this document:

1. DATASET PART I:

Description of the existing dataset containing audio- or video-recordings and transcriptions of job interviews (collected from 2011-2018). This dataset contains many job interviews and serves as a corpus of data to describe **benchmark cases** of identity work in job interviews in general.

2. DATASET PART II:

Description of the new dataset containing video-recordings of job interviews (collected from 2019-2023). For each of these job interviews, we collect additional data that is directly linked to these job interviews:

- a. Coded job interviews: the job interviews will be coded for specific linguistic elements.
- b. SRM-interviews: these are interviews in which we watch clips of the video-recordings of the job interviews together with the recruiters to get their comments on the interactions in the job interviews.
- c. Questionnaires: for each job interview, the recruiter fills out a questionnaire where he or she describes the context for the job interview in terms of War for Talent vs. War for Jobs.

This dataset is **specifically compiled for** this research project and serves as a corpus of data to describe identity work in **War for Talent**-job interviews.

3. DATASET PART III:

Description of the set of ethnographic interviews: this set of data is not linked to concrete job interviews but contains interviews with recruiters where they describe the **recruitment context of the War for Talent in general**.

4. DATASET PART IV:

Questionnaires collected during experiment: this final element in the data contains the questionnaires that will be filled out by master students in academic year 2023-2024 during an experiment.

### 4.2. Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.

The report described above will provide a description of all data and metadata created. To accompany this report, we will create an overview (.xlsx) that indicates for each job interview in the dataset which type of metadata is available. This document will make clear for which job interviews specific types of metadata have been collected and created: this differs of course a great deal between the existing corpus and the new corpus collected specifically for this project. This overview will be kept up to date throughout the project, as in each phase new types of data and metadata will be collected/created.

## **5. DATA STORAGE AND BACKUP DURING THE C1-C2 PROJECT**

### **5.1. Where will the data be stored?**

On KU Leuven OneDrive for Business (used in combination with KU Leuven authenticator) and the hard drive of the PhD-student's laptop (which is bitlocker protected)

### **5.2. How will the data be backed up?**

The use of OneDrive for Business provides backup. Also, two external hard drives (bitlocked) are used as back-up.

### **5.3. Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.**

KU Leuven's OneDrive for Business provides researchers with 2TB of storage, which should suffice for this project.

### **5.4. What are the expected costs for data storage and backup during the project? How will these costs be covered?**

With the C1-project funds, two external hard drives have been acquired for back-up.

### **5.5. Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?**

OneDrive for Business ensures safe storage in combination with KU Leuven Authenticator. Only the PI and PhD-student have direct access to the authentic data.

## **6. DATA PRESERVATION AFTER THE END OF THE C1-C2 PROJECT**

### **6.1. Which data will be retained for the expected 10 year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).**

The entire dataset will be preserved after the end of the project for at least 10 years.

### **6.2. Where will these data be archived (= stored for the long term)?**

The PI will preserve the entire dataset at least 10 years after completion of the project on KU Leuven OneDrive for Business conform the KU Leuven RDM policy. It can however, due to the third party agreement, not be shared with external researchers.

### **6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?**

KU Leuven OneDrive for Business does not require any costs for the researchers.

## **7. DATA SHARING AND RE-USE**

**7.1. Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions or because of IP potential)?**

Yes, since this data is highly privacy sensitive. As stipulated in the informed consent forms that the participants sign, the data cannot be shared with anyone other than the project collaborators.

**7.2. Which data will be made available after the end of the project?**

NA

**7.3. Where/how will the data be made available for reuse?**

NA

**7.4. When will the data be made available?**

NA

**7.5. Who will be able to access the data and under what conditions?**

NA

**7.6. What are the expected costs for data sharing? How will these costs be covered?**

NA

## **8. RESPONSIBILITIES**

**8.1. Who will be responsible for the data documentation & metadata?**

Dorien Van De Mieroop and Melina De Dijn

**8.2. Who will be responsible for data storage & back up during the project?**

Dorien Van De Mieroop and Melina De Dijn

**8.3. Who will be responsible for ensuring data preservation and sharing?**

Dorien Van De Mieroop and Melina De Dijn

**8.4. Who bears the end responsibility for updating & implementing this DMP?**

The end responsibility for updating and implementing the DMP is with the PI/supervisor, Dorien Van De Mieroop.