

Claudia Mejía de la Pava

University of Salamanca
Spain
clm.delapava@usal.es

UDK 81'33:303.42

DOI: 10.4312/vestnik.16.155-174

Izvirni znanstveni članek



Silvia Manessi

University of Salamanca
Spain
silvia.manessi@gmail.com

HANDS-ON DATA GATHERING: VIABILITY OF RESEARCH METHODOLOGIES IN TWO LINGUISTIC STUDIES

1 INTRODUCTION

This paper aims to present different methodological approaches in eliciting sociolinguistic data vis-à-vis the research questions of a given study and its scope. To do so, it highlights the context and aims of the two sociolinguistic studies as examples of such linguistic approaches. The first study is conducted within the field of hang gliding and the English used by some of the members of this community. The second one is centred in the context of the management of staff working for European Institutions and their being acquainted with the concepts of equality and diversity.

Choosing the most effective research method in sociolinguistic analysis is of the utmost importance for the outcome of the research itself.

Qualitative research seeks to understand the ways people experience events, places, and processes differently as part of a fluid reality, a reality constructed through multiple interpretations and filtered through multiple frames of reference and systems of meaning-making. (McGuick 2016, 10)

While, at a first glance, the two studies seem far apart they still portray common elements, such as the analysis of the language used in their community of practice, which allows the opportunity to assess, compare and evaluate the adequacy of the different data elicitation methods in order to identify their suitability for each of the two studies and the lessons learned from these investigations.

The first research presented in this paper explores the use of spoken English as a second language (L2), among the members of the International Hang Gliding Community. This study of the interactions within a specific multicultural community of practice (CoP), similar to studies in different CoPs (Cogo and Dewey 2012; Ehrenreich 2009; Ranta 2009), also considers the fact that the interactions occur in both general and specialized contexts. The intention is to observe the production of natural spoken language and gain insight into the peculiarities of this variety of English, focusing on morphological, semantic and syntactic features, as well as code-switching, pragmatics and linguistic variation. A better understanding of the use and production of this specific variety of English within a specific CoP could contribute to areas such as English as a Lingua Franca (ELF), Multicultural Communication, English for Special Purposes (ESP) and/or Specialized Translation.

In short, it can be said that this CoP is composed of a heterogeneous group of individuals from different nationalities and linguistic backgrounds, belonging to various age groups. They have diverse professions, educational and English proficiency levels, as well as varying degrees of athletic skills, and have been involved in the sport, and its community, for different periods of time, ranging from a couple of years to several decades. Nonetheless, when they come together to participate in international events they communicate in English with each other. It is the possibility of studying the peculiar variety of spoken language produced by these athletes which appeals to the researcher.

In order to conduct this exploration, it was necessary to have access to suitable material to analyse, which led to the decision to self-gather spoken data with the purpose of creating our own specialized corpus to conduct the research.

The second study included in this paper focused on the degree of awareness of equality and diversity principles among staff working for the EU institutions, given their employment regulatory framework, which is composed of the EU Staff Regulations in conjunction with the Condition of Employment of Other Servants of the European Union, and by their implementing provisions. While the CoP in this study is quite different from that in the first study, the use of the English language as an element of identity of the different participants in the CoP is common to them both (Gee 2011:106).

The aim of the second study was to understand, by way of a written questionnaire, how the staff rules are used, on a daily basis, by a community of HR and Legal Officers working for the European institutions, and whether these rules foster a message on equality and diversity by using language that could substantiate and support those principles.

Further into the investigation, the methodology developed around the need to understand how the EU Human Resource Management (HRM) Legal Framework may have an impact on the creation of the identity of the staff members working for the different EU institutions in general, but also particularly vis-à-vis the values of equality and diversity. The methodology used to answer this second research question consisted of the analysis of the discourse around the EU HRM Legal Framework, in combination with the analysis of the replies received from a questionnaire distributed among EU staff members.

2 DESCRIPTION OF THE COMMUNITIES OF PRACTICE

The idea of Communities of Practice (CoP), developed in the early 1990s by Lave and Wenger, was first transferred into the field of sociolinguistics in 1992 by Eckert and McConnell-Ginet (Piller 2012). The three main aspects of a CoP are “mutual engagement, joint enterprise and shared repertoire” (Culver and Trudel 2008). Piller further expands the description as an assemblage of individuals bound by their commitment to an activity, resulting in a series of practices emerging from this shared venture. It diverges from a traditional community since a CoP is defined concurrently by its affiliation and the practices deriving directly from the affiliation. Each study presented in this paper is concerned with multinational and multicultural CoPs, one pertaining to the field of sports and the other to the professional field.

The International Hang Gliding Community (IHGC), besides being multicultural and diverse, also embodies a dynamic and cohesive group within the domain of sports, more specifically in the field of recreational aviation. Every year its members come in contact with fellow pilots from many different countries during a series of international competitions that take place in various regions of the world. With events ranging from four to five days up to two weeks, its members are in continuous interaction, both on an athletic as well as a personal level. Having observed several of these interactions and recognizing the complexities involved, the researcher was able to identify the potential for a field study of their naturally occurring, non-academic, generic as well as specialized spoken language.

With regard to the second study object of this paper, the CoP sample includes Human Resources and Legal Officers working for EU institutions, who are at the same time the subject of the focal legal framework, as are all other staff working for EU institutions, and have direct responsibility for reading and understanding the framework, given their deep knowledge and expertise on the subject matter. HR Officers in the EU institutions are, in fact, also particularly involved in translating the different aspects included in the legal framework into plain language. Each EU institution and each agency has internal decisions, notes, explanations and FAQs on the provisions applicable to EU civil servants, and this internal material is often prepared by HR and Legal Officers. Staff can access this broad range of material, but their exposure to primary legislation might in practice be rather limited. For this reason, it seems sensible that the population of HR and Legal Officers is the most relevant one as the subject of this research.

For the analysis of the discourse on equality and diversity in the workplace, the EU institutions represent a unique territory for linguistic research. Therefore, the aim of this study was to highlight the uniqueness of the discourse around equality and diversity and understand the identity of these staff members, *vis-à-vis* this topic, in their own CoP.

As mentioned in the introduction of this paper, the research projects presented in this work might seem quite different at a first glance. However, after a closer look it is

possible to find several similarities that offer the common ground on which we are basing our parallel analysis of the type of data-gathering methodologies implemented. Table 1, below, summarizes some of the general traits that define the two population groups in this project (IHGC is the International Hang Gliding Community, and MIEU is the Members of the Institutions of the European Union):

Table 1: General characteristics of the sample populations (by authors)

	IHGC	MIEU
Universal Population	Approx. 12,700	Approx. 12,000
Sampling Frame (SF)	Approx. 2,000	Approx. 600
Sample Population (SP)	116 interviewees (89 in-person/ 27 virtual)	53 respondents
Native Language - SP	Catalan, Czech, Danish, Dutch, French, German, Greek, Hungarian, Italian, Japanese, Macedonian, Norwegian, Polish, Portuguese, Russian, Slovene, Spanish, Swedish, Turkish and Ukrainian.	Bulgarian, Croatian, Czech, Dutch, English, Finnish, French, German, Greek, Hungarian, Italian, Lithuanian, Maltese, Polish, Romanian, Slovene, Spanish, and Swedish.
Education Level - SP	Secondary School: 33 Undergraduate / Graduate: 82	Secondary School: 0 Undergraduate / Graduate: 53
English Proficiency - SP	A Level: 28 B Level: 46 C Level: 42	A Level: 0 B Level: 2 C Level: 51

It is possible to see that while the size of both universal populations may be similar, the sampling frame (SF) size is quite different, with the SF for the IHGC being almost 3.3 times larger than that for the MIEU. Nevertheless, the pool of 116 interviewees comprising the sample population (SP) of the IHGC represent 6% of the SF, while the 53 respondents that constitute the SP of the EU institutions represent 9% of that specific SF, making both sample populations relatively similar. The linguistic scope of both studies comprises a similar number of languages (20-IHGC / 18-MIEU), mostly European (Czech, Dutch, French, German, Greek, Hungarian, Italian, Polish, Slovene, Spanish and Swedish). The major difference between both sets of languages is the exclusion of English and the inclusion of Japanese for the IHGC and the inverse situation for the MIEU. However, it is important to note that this dissimilarity is intrinsic to the nature of each study. In the former case, the study focuses on users of English as a second language, and therefore, members of this community with English as their first language were filtered out. In the latter case, the study centres solely on the institutions within the EU; therefore, while English is included, Japanese was never taken into consideration since it is not a language of an EU member state.

Looking at the education and English proficiency levels of the SPs, it is possible to notice a difference in the composition of both groups. Whereas in the SP of the IHGC 28% of the individuals completed secondary school and 71% have either an undergraduate or a graduate degree, in the SP of the MIEU no participants fall in the former category and 100% of the individuals belong to the second one. There is also an outlier within the IHGC SP, represented by a person who did not complete their secondary school education.

As far as English proficiency levels are concerned, we may observe that most of the individuals (40%) have an intermediate (B) level of proficiency in the SP of the IHGC, followed by a 36% having an advanced (C) level, and 24% with a basic (A) knowledge of English. This is in contrast with the MIEU SP, where 96% of the participants have an advanced knowledge of the language and the other 4% have an intermediate level of proficiency, with no one in the basic knowledge category. It is significant to state that the levels in both categories within the MIEU SP may be accounted for by the education requirements in place in order to be able to work within the EU institutions.

Figure 1, below, shows the age distribution of the participants of both groups:

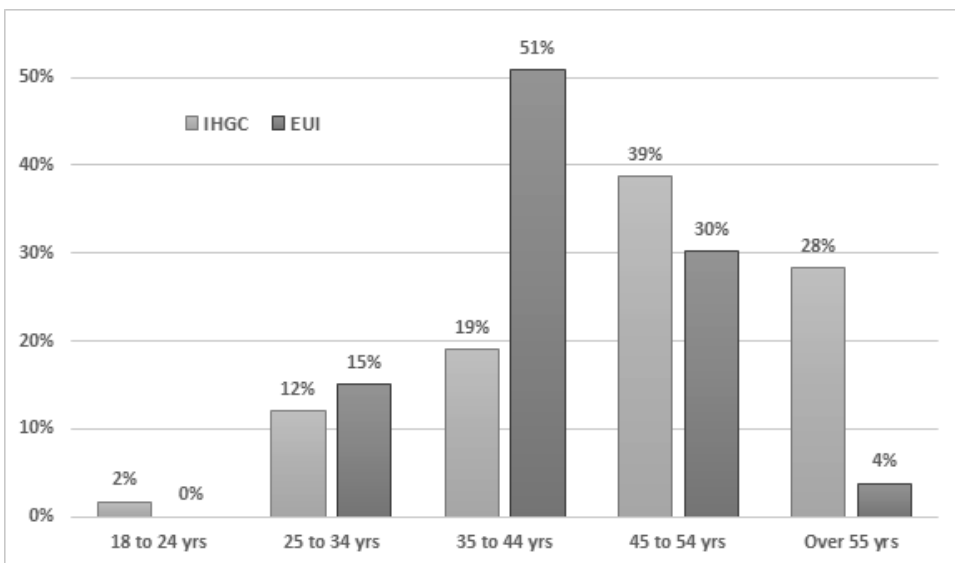


Figure 1: Age distribution of the sample populations (by authors)

It is possible to observe that both groups have a similar age distribution as far as three of the age brackets are concerned: (a) less than 10% of participants are between the ages of 18 and 24, (b) more than 10% but less than 20% are between the ages of 25 to 34, and (c) from 30% to 40% are between the ages of 45 to 44. Nevertheless, it is also possible to observe several variances between the groups. On the one hand, the largest

group for the IHGC is represented by those individuals between the ages of 45 to 54 (39%), while the 51% of the participants within the MIEU group are between the ages of 35 to 44. On the other hand, we have inversely divergent tendencies as far as the middle and last age brackets are concerned: 19% of the IHGC SP are between the ages of 35 and 44, while 51% of the MIEU SP belong to this same bracket, which is more than twice the number of individuals. The opposite situation can be found within the last bracket, over 55, where we see 28% of the IHGC participants and only 4% of the MIEU, indicating another substantial difference between the two groups.

Finally, we can observe a similar situation in the distribution of the years of involvement in each group, as seen in Figure 2, below, showing the members of the sample population of the IHGC in green, and MIEU staff in blue.

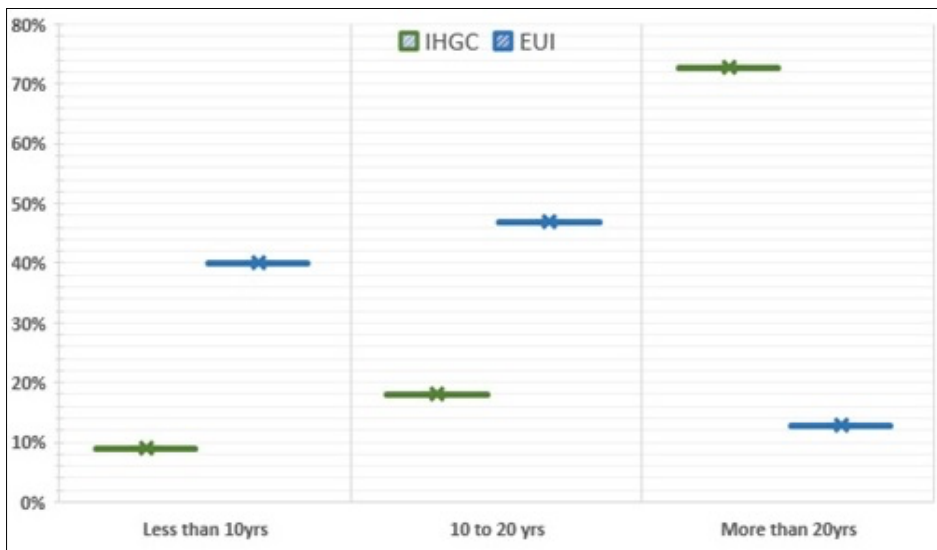


Figure 2: Distribution of years of involvement of sample populations (by authors)

The trait which can be said to be the most noticeable is the fact that while the vast majority of the hang glider pilots (73%) have been involved in the sport for over 20 years, only 13% of the staff members within the MIEU have been involved in these institutions for over 20 years. Moreover, a second notable characteristic is that only 9% of the IHGC group has been practicing the sport for less than 10 years, while this same bracket corresponds to 40% of the MIEU sample, an indication that this particular tendency in both groups is inversely related.

3 LITERATURE REVIEW

In order to outline the process and prepare the material required to obtain the specialized corpus for our first study, the researcher based the methodology on several publications that focus on works related to research methods in the fields of linguistics (O’Keeffe & McCarthy 2012; Podesva & Sharma 2013; Williamson 2018), corpus and applied linguistics (Hall, Smith & Wicaksono 2011; Kubler & Zinmeister 2015; Weisser 2016), as well as qualitative approaches, in general, and research methodology in social studies research (Gobo 2004; Lewis & Ritchie 2003; Neuman 2014).

Additionally, it was deemed essential to complement the research work by looking into studies concerning the different topics this project embraces, such as those encompassed in Paltridge and Starfield’s volume on *English for Specific Purposes* (2012), works on English as a Lingua Franca (Cogo & Dewey 2012; Mauranen 2007; Mauranen & Ranta 2009; Pickering 2006; Seidlhofer 2001 & 2005), as well as studies pertaining language teaching and translation correlated to corpus linguistics (Bernardini 2003; de Jesus 2014; Tagnin & Murakami 2014).

With regard to the studies on which the second research is based, a questionnaire for HR and Legal Officers has been developed to elicit direct data from this CoP (Brace 2004; Krug & Sell 2013; McGuirk 2016; Wiersma 2011).

To enhance the analysis of the language used in the EU HRM Legal Corpus, the study touched upon different features proposed by several of the aforementioned studies, in particular taking into account studies related to discourse in the workplace (Bhatia 1993; Grant and Iedema 2004; Danson, Green and McQuaid 2005; Cotter and Marschall 2006; Rodríguez 2006; Mayr 2008 and 2015; Shena et al., 2009; Koester 2010; Ten Thije and Maier 2012; Kärreman 2014; Jackson 2014; Holmes 2015).

Furthermore, the analysis of the discourse of the EU HRM Legal Corpus was based on research related to the discourse on equality and diversity (Holmes and Meyerhoff 2003; Hummert Garstka, Ryan and Bonnesen 2004; Baker 2005 and 2018; Ahmed 2007; Mautner 2007; Harrington et al., 2008; Goddard and Mean 2009; Talbot 2010; Özbilgin and Tatli 2011; Tatli 2011; De Graaff and van der Wal 2011; Altan et al., 2014; Grue 2014; Hord 2016; Sardar 2018).

4 RESEARCH HYPOTHESES

The questions driving the first study focus on looking for evidence to establish, first of all, whether occurrences such as semantic and syntactic deviations, among others, are related entirely to the speakers’ first language (L1) or if indications of a third, neutral language (LN) or a different factor can be observed. Secondly, to assess whether the peculiarities of this language variety appear to be linked mostly to the use of specialized vocabulary or

if they can be found within the register of non-specialized language as well, and to what extent. In the third place, to assess whether it is possible to recognize a clear group identity embedded in this discourse or the presence of a sociolect among the data collected, and whether it is feasible to identify how gender is represented within this community. This study will also analyse the data in an attempt to find out if the differences identified within the discourse appear to hinder the communication process among group members, as well as if there is a need to implement measures aimed at improving the oral interactions within the population.

In order to tackle the important and controversial matter of the second study, a fundamental research question has been identified as key for this paper, i.e. the need to establish what the EU HRM Legal Framework's discourse is inside the different EU institutions, and the degree to which it delivers a message of universal values of equality and diversity. The methodology used to address this research question was conducted in two dimensions: on the one hand, by analysing the semantic and syntactic features of a corpus of documents including all the HRM-related provisions common to all European institutions and, on the other hand by looking into the findings of the corpus linguistics analysis and the analysis of the discourse of this corpus of documents, together with the comparison of its narrative against other similar genres.

It is a contention of the authors of this paper that the methodological framework and research methods of the studies being considered align with each of the modalities of data gathering employed. The authors' decision of implementing a certain method of data elicitation (oral vs written) was determined by their own research interests (Jones 2013). Oral interviews were deemed appropriate for the first study, after establishing that the analysis would focus on linguistic features within the spoken language used by the community of hang gliding athletes. By the same token, for the second study a written questionnaire was considered a suitable method to facilitate this task, since it would target the influence of a set of written norms in the context of an international working environment.

We thus start with the context of the studies and the linguistic features that will be analysed, moving through the two CoPs as unique places of interaction and identity, to finally reach the aim of capturing the data itself, in order to obtain a description of the linguistic features and proceed to the analysis and/or to seek improvements by suggesting adjustments and refinements to the language being used.

5 METHODOLOGY

Both studies, considering the scope and nature of their subjects, entailed different methods to collect their data, which are described below. For one study, we present information regarding the design and implementation of a series of interviews – mostly in-person

and others virtually – while focusing here on the latter, which took place between 2018 and 2020. For the second study, we provide details regarding the design and implementation of an online questionnaire conducted throughout April and May 2019.

For each study, the researchers opted for what seemed to be the most appropriate method to elicit the data. For the first study, its focus on spoken language, combined with the ethnographical emphasis, were the main factors driving the decision as to which method to adopt. For the second study, a different approach was considered vis-à-vis the linguistic features of interest, with these being used for identity building within the CoP (Bathia 1993; Gee 2011).

With the extent and objectives of the first study in mind, and after weighing different options, the researcher decided to use recorded in-person interviews with members of the International Hang Gliding Community (IHGC) as the main data-collection method. That being said, it is relevant to mention that, due to travel restrictions and lockdowns caused by the COVID-19 pandemic, the researcher was compelled to adapt to the changed circumstances and implement a virtual interviewing protocol in order to fulfil the sampling quota. This section will focus mostly on the 89 in-person interviews, out of the total of 116, which were conducted during the researcher's visits to eight hang gliding competition venues in six countries, and 10 meetings in eight different countries.

The main advantages of opting for this kind of interview include the opportunity to collect natural speech and a diversity of registers with similar content, thus enabling a comparative analysis, while also ensuring that those individuals volunteering their time would feel at ease with the process.

There are three phases to the approach used in this study: (a) the design of the structured questionnaire, (b) the logistics involved in meeting with the participants, and (c) how these interview sessions were conducted.

With the right structure, the questions used would allow the elicitation of the different registers used by the members of this community: casual or informal conversation, technical or specialized vocabulary and evocative language. This led to the decision to use only open-ended questions to avoid any binary yes/no answers (Legard et al., 2003), as well as any short answers such as “maybe”, “I don’t know”, etc. By having the same questions for every participant, the researcher aimed at collecting data that would be homogeneous in subject yet possibly diverse in structure, facilitating a comparative analysis. Lastly, by designing questions flowing from general to more specific topics, related to known subjects while, at the same time, keeping them relatively simple to avoid intimidating the interviewees, the researcher could improve the probability of collecting higher quality data of a conversational nature (Arthur and Nazroo 2003).

The result was a series of seven questions covering both areas of general and specialized knowledge, related to hang gliding as well as providing the opportunity for the interviewees to share some of their individual experiences, in order to produce some evocative content.

Besides the seven questions mentioned above, given that the implementation of these interviews entailed a one-on-one interaction, where the participants were recorded while speaking, it was necessary to consider the ethical aspect of the process. Therefore, the researcher needed to make sure each interviewee would be properly informed of the process, its objective and how the data would be utilized, while verifying that their participation was categorically voluntary. Furthermore, it was necessary to get an authorization stating that the interviewees fully understood what they were agreeing to, and all of this needed to be done in such a way to establish a trusting environment and a collaborative relationship (Eckert 2013).

To satisfy these requirements, the research was based on a two-step format. A privacy and consent form with the introduction of the research was created, explaining how the data would be handled and requesting contact information for any follow-up that was needed. The presentation of this information and document, together with the collection of the signed consent form, was conducted directly during the initial stage of each interview.

The last stage of the interview sessions – after the recording – was the collection of certain personal information. In ethnographic research, many scholars consider that having access to personal data from the pool of interviewees represents an opportunity to gain better insights into the subject of interest (Legard et al., 2013), which could be helpful in better understanding the community or the features observed in their speech during the analysis phase. The researcher thus created a second PDF document that was filled out with help of the interviewee after the last question had been completed. Collecting this information after the interviewee had answered all the questions, and not before, was done in order to avoid a sense of apprehension or self-awareness with regard to the topics being discussed (Schilling 2013), which could – consciously or subconsciously – impact their mood or disposition to respond to the questions.

The potential volunteers were identified within the sampling frame obtained for this research, and those who participated in the study had 31 different nationalities and 20 different linguistic backgrounds. They were all hang gliding pilots who needed to use English, as a lingua franca, when participating in hang gliding events and interacting with staff members, officials and their peers.

The best course of action to optimize the implementation of the interviews appeared to be visiting competitions and conducting them on site, as well as attempting to meet with volunteers during the journeys to and from the venues, or amidst work-related travel. This involved extensive planning as well as understanding the importance of taking the potential participants' time constraints and engagements into consideration. These concerns influenced the overall shape of the interview itself, such as the length of the sessions, in order to avoid speaker fatigue (Chelliah 2013), and since all the participants were on a similar tight daily schedule, the researcher had an obligation to be mindful of their time (Levon 2013). Considering all the different factors, it was decided that each session should range between 45 minutes to an hour.

Most of the 116 interviews (79) were conducted during hang gliding events and under more challenging conditions compared to the other 10 face-to-face interviews conducted outside a competition setting, or the 27 virtual interviews. With regard to those interviews held during competitions, we can begin with the challenge presented by scheduling them. First of all, it was not possible for the researcher to plan these interviews ahead of time. Secondly, during any competition the researcher had to work around the participants' tight schedules, and generally there was only a short window in the morning, before the beginning of each race, and another small window in the evening, after the flights. On average, the researcher was only able to conduct between one to four interviews per day. Another challenge was the locations, as hang gliding is an outdoor sport and the events take place in remote settings, mostly due to the need for ample space both on the ground and in the air to provide a suitable and safe racing arena. This ruled out having access to places such as meeting rooms, classrooms, offices or quiet indoor spaces, and thus there was inevitably some background noise captured by the recording equipment (Podesva and Zsiga 2013).

The focus of the second study highlighted in this paper was the language and EU staff's understanding with regard to the values of equality and diversity (Brace 2004; McGuirk 2016). To complement the corpus linguistic and discourse analysis of the EU Staff Regulations, a questionnaire for HR and Legal Officers working for the EU institutions was designed, with the aim to gather meaningful data for the people who must work with this legal framework.

An important criterion in the development of the survey was the mode of submission of the questionnaire for HR and Legal Officers working for the different EU institutions, as these institutions are present in all of the Member States, and thus the target respondents were located throughout the EU.

For this reason, the most logical approach was to circulate the questionnaire by using an online survey tool. The added value of this was the possibility to reach a relatively large population of participants with a short and defined period of time. Another advantage was that the format gave more freedom to the participants taking part in the survey, and allowed the analysis of the resulting data in a timely manner.

One more added value of administering online surveys was the ease of accessing the questionnaire via computer or mobile phone, which was deemed particularly important since the participants were located in different cities across Europe. With the use of Lime Survey each person could access the questionnaire via a link, and they were also asked to help increase the amount of data collected by sharing this link with their personal contacts who have HR and legal backgrounds.

Lime Survey¹ is a free-to-access licensed tool for the design of online surveys, where the author can adapt the open-source code to the specific needs of their

1 <https://www.limesurvey.org/>

research. Its features include the option of having an unlimited number of surveys, questions and participants, a broad selection of pre-established questions (a total of twenty-eight types) and the possibility to avoid including any advertisements. A major advantage of Lime Survey is that it supports more than 80 different languages, and it includes a design editor that makes it possible to change and define the structure and layout of the survey, with a relatively simple layout being used in this work. Moreover, since it is an open-source tool, the system can be installed on the researcher's own computer, guaranteeing full control over the data acquired. Reaching out to the population was also convenient and easy, given the many internal channels for the distribution of information within EU institutions.

Another criterion that was considered in this study was minimizing the effort needed by the respondents to complete the survey (Krug and Sell 2013). In order to achieve this, all the questions and statements that were to be answered using a Likert scale were designed to be as concise and short as possible.

Another important factor when designing the questionnaire relates to clustering questions and statements in specific areas. The survey was thus structured using different thematic areas, to help the participants go into greater depth with regard to the different topics. It is important to underline that all parts of the survey were made fully accessible to the respondents at any time, and they could freely go back and forth throughout the questionnaire before submitting it.

The different parts of the questionnaire were designed to guide the participants to its completion, starting from broader questions and statements and then moving on to more specific ones. The aim was two-fold: on the one hand, to make the participants' journey through the survey easy and smooth, increasing the possibility of a high completion rate, and, at the same time, help them in dealing with the most sensitive or controversial questions by offering them the more general and less controversial items first.

6 RESEARCH FINDINGS

Despite the complexity of the interview process in the first study, which included extensive traveling and challenging conditions, it is possible to say that all this work resulted in valuable material and fulfilled the objective set by the researcher, with a total of 17.5 recorded hours of spoken data being obtained. Moreover, by selecting an oral, in-person interview as her main method, the researcher was also able to benefit from the inherent opportunity for her to become an observer, as well as sometimes acting as a participant-observer in some of the other activities within the community. This enriched the quality of the research, by allowing her, beyond the interview sessions, to better understand the group she was aiming to describe (Geznuk 2003), as well as broadening her personal experience.

During the initial analysis of the material that was obtained, it was possible to identify an extensive range of linguistic phenomena, falling within the scope of the initial research questions, which the researcher then went on to further analyse in the following stage of the study. These phenomena include various instances of coinages and calques, code-switching into the speakers' first language and some of examples of code-switching in a third language (neither English nor the speaker's mother tongue). The corpus obtained through the data-gathering and processing phases contains numerous linguistic structures, which may prove significant for the data analysis phase and the study's foci. Furthermore, the data offered valuable insights on the use of gender language within this community. In addition to the expected wide presence of gender pronouns within both the specialized and generic vocabulary, the data also seem to point towards a much higher than expected presence of gender-neutral speech. The preliminary findings reached by analysing the answer to one specific question, which gathered the most suitable material to analyse this feature, show that the vast majority of the interviewees (90%) used gender-neutral language, 9% used a mixture of gender-neutral and gender-specific language, and 1% explicitly used gender-inclusive language.

Finally, it seems plausible to say that this study may complement current lines of research, such as those mentioned in the literary review above, since this project combines different aspects of other studies, such as: oral language produced by users of English as their L2 and/or a lingua franca – gathered in a natural setting – that includes a certain amount of specialized vocabulary, yet remains outside the academic or work settings.

With regard to the second study, the research tried to complement the data obtained from the corpus linguistics and discourse analysis of the HRM Legal Corpus with the data elicited from the questionnaire.

The specific fields of research in the second study were related to four main areas: gender mainstreaming, the rights of LGBTI persons, persons with a disability and the inclusion of elderly staff (Holmes and Meyerhoff 2003; Hummert et al., 2004; Baker 2005 and 2018; Ahmed 2007; Mautner 2007; Harrington et al., 2008; Goddard and Mean 2009; Talbot 2010; Özbilgin and Tatli 2011; Tatli 2011; De Graaff and van der Wal 2011; Altan et al., 2014; Grue 2014; Hord 2016; Sardar 2018). The data related to gender mainstreaming in particular showed some unexpected and interesting results.

Contrary to the other study, in which gender-biased language was less frequent, the corpus and discourse analysis conducted for this study showed many examples where the neutrality of gender was not respected in the language used. Although in equality and diversity policies gender is usually seen as a very significant issue, due to the political and sociological discourse in recent years only a minority of the respondents to the survey felt that there was any inconsistency in the use of gender pronouns within EU institutions.

These responses contradicted the data of the corpus linguistics analysis, which highlighted the predominant use of gender-biased language, showing a more comprehensive and sensitive understanding of gender mainstreaming. This finding is also related to the

fact that current EU gender policies specifically target women in managerial positions.

With regard to the other categories that were investigated, the results of the survey showed an increased number of neutral responses (3 on a Likert scale from 0 to 5), showing that respondents to the questionnaire seemed to have less sensitivity with regard to the inclusion of LGBTI staff.

Similarly, almost two-thirds of the respondents to the questions related to persons with a disability offered rather neutral responses with regard to the language used in the EU HRM Legal Framework. This could be due to there being less exposure to the rights of persons with a disability and a rather undeveloped sensitivity with regard to this. This last hypothesis could be confirmed by the fact that two-thirds of the respondents stated that the use of the term “disabled” does not have any biased connotations.

With regard to the rights of the elderly, more than half of the respondents gave neutral responses (3 on a Likert scale from 0 to 5), possibly for the same reasons indicated above.

7 CONCLUSIONS

Each of the two studies described above were effective in the analysis of the linguistic features of their communities of practice, and the two different methodologies used both added value and problems to their respective research projects, as highlighted below.

Looking back at the method used with the hang gliding community, the in-person interviews allowed for a meaningful amount of data to be collected, and also provided the opportunity for the researcher to have better interactions with the population sample and the community overall, including the possibility of complementing the study by adding an ethnographic approach to it.

However, this method did present several challenges, such as the complexity of the logistics and the execution of the interviews in the remote areas where hang gliding competitions take place. In addition, while attending such events did provide an excellent opportunity for recruiting participants, it also created time constraints that resulted in relatively short interviews and an inability to fully control the sound quality of the recorded conversations. Overall, however, weighing the advantages and disadvantages identified by the researcher, the results show that the time and effort invested in this approach were definitely worth it.

With regard to the research conducted in the European Union context, the use of an online questionnaire proved to be highly effective vis-à-vis the aim of the data collection. While this approach (using Lime Survey) proved to be quite easy, the decision to make use of a written questionnaire to elicit data could potentially bring some complications in the design phase, as once the questions are completed there is no way to change them in order to increase the comparability of the data, if needed. Much effort thus needs to be put into designing the questionnaire before it is used, so that all the data needed are collected and that there is no missing information that is important for the research, and “the

information gathered is only as good as the questions that are asked” (Brace 2004, 1).

If such a questionnaire is well targeted and designed, though, an added value compared to oral interviews is the possibility for the respondents to reflect on the questions and topics, and so provide more meaningful insights than in an oral interview.

This proved to be the case for questions related to the improvement of the language of the EU Staff Regulations, which were one focus of the overall research. The respondents had the time to reflect on the different issues that were highlighted in the questionnaire and the Likert-scale questions, as well as having the possibility to include their thoughts and suggestions for future reference.

As a general lesson learned from both studies, it should be mentioned that the time and effort invested in eliciting direct data from the focal communities of practice were important in disentangling the language used and thus better addressing the research questions. If we had decided not to pursue direct data gathering with interviews and questionnaires, but only based the research on already available material (such as, for example, the Corpus of EU HR related rules), both studies would have been utterly incomplete (Brace 2004; Krug and Sell 2013).

At the same time, while acknowledging the balance between effectiveness and time needed, vis-à-vis both the data gathering and the respondents, a second round of questions could have helped to deepen the research, in particular regarding a few aspects of future developments with regard to the interactions and identity of the community of practices themselves (Brace 2004; Krug and Sell 2013; McGuirk 2016; Wiersma 2011).

While we are convinced that the modes of data gathering used – i.e., oral interviews for linguistic analysis mostly related to the use of spoken language and a written questionnaire to examine the discourse and identity around a written legal framework – were successful for both studies, the main takeaway from the experience of both projects was related to the meaningfulness of the direct interaction with the participants of both communities of practice, which can be said to have made the aims and evolutions of the studies more significant and relevant.

BIBLIOGRAPHY

- AHMED, Sara (2007) The language of diversity. *Ethnic and Racial Studies* 30, 235–256. DOI: 10.1080/01419870601143927.
- ALTAN, Levent et al. (2014) *Implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation*. Luxembourg: Publications Office of the European Union.
- ARTHUR, Sue/James NAZROO (2003) Designing Fieldwork Strategies and Materials. J. Ritchie/J. Lewis, *Qualitative Research Practice*. London/Thousand Oaks/New Delhi: Sage Publications, 109–137.
- BAKER, Paul (2005) *Public Discourses of Gay Men*. London: Routledge.

- BAKER, Paul (2018) Language, sexuality and corpus linguistics. H. Motschenbacher (ed.), *Journal of Language and Sexuality: Taking stock and looking ahead* 7/2, 263–279.
- BERNARDINI, Silvia (2003) Designing a Corpus for Translation and Language Teaching: The CEXI Experience. *TESOL Quarterly* 37/3, 528–537.
- BHATIA, Vijay (1993) *Analyzing genre: language use in professional setting*. London: Longman.
- BRACE, Ian (2004) *Questionnaire design: how to plan, structure, and write survey material for effective market research*. London: Sterling.
- Center for Multilingual, Multicultural Research (2003) *Occasional Papers Series*. Los Angeles: Rossier School of Education, University of Southern California.
- CHELLIAH, Shobhana (2013) Ethnography and Recording Interaction. R. Podesva/D. Sharma (eds.), *Research Methods in Linguistics*. New York: Cambridge United Press, 51–73.
- COGO, Alessia/Martin DEWEY (2012) *Analysing English as a Lingua Franca: A Corpus-Driven Investigation*. London: Continuum.
- COTTER, Colleen/Daniel MARSCHALL (2006) The persistence of workplace ideology and identity across communicative contexts. *Journal of Applied Linguistics and Professional Practice* 3/1, 1–24. <https://doi.org/10.1558/japl.v3i1.1>.
- CULVER, Diane/Pierre TRUDEL (2008) Clarifying the Concept of Communities of Practice in Sport. *International Journal of Sports Science and Coaching* 3/1, 1–10. DOI: 10.1260/174795408784089441.
- DANSON, Mike/Anne GREEN/Ronald MCQUAID (2005) Introducing employability. *Urban Studies Journal Limited* 42/2, 191–195. DOI: 10.1080/0042098042000316092.
- DE GRAAFF, Maaike/Ronald VAN DER WAL (2011) *On the shoulder of the predecessors: striving for diversity in the Dutch police*. The Hague: EDPOL.
- DE JESUS, Silvana (2014) O corpus combinado e a pesquisa nos Estudos da tradução baseados em corpora. *Letras & Letras* 30/2, 291–321. DOI: 10.14393/LL60-v30n2a2014-14.
- ECKERT, Penelope (2013) Ethics in Linguistic Research. R. Podesva/D. Sharma (eds.), *Research Methods in Linguistics*. New York: Cambridge United Press, 11–26.
- EHRENREICH, Susanne (2009) English as a Lingua Franca in Multinational Corporations: Exploring Business Communities of Practice. A. Mauranen/E. Ranta (eds.), *English as a Lingua Franca: Studies and Findings*. Newcastle: Cambridge Scholars, 126–151.
- GEE, James Paul (2011) *An introduction to Discourse Analysis*. New York: Routledge.
- GENZUK, Michael (2003) A Synthesis of Ethnographic Research. Rossier School of Education of Southern California – Center for Multilingual, Multicultural Research (ed.), *Occasional Papers Series*. 1–10. [Accessed on 1 March 2024]. https://web-app.usc.edu/web/rossier//33/Ethnographic_Research.pdf.
- GOBO, Gianpietro (2004) Sampling, Representativeness and Generalizability. C. Seale/G. Gobo/J.F. Gubrium/D. Silverman (eds.), *Qualitative Research Practice*. London: SAGE, 405–426.
- GODDARD Angela/Lindsey MEAN (2009) *Language and gender*. New York: Routledge.

- GRANT, David/Rick IEDEMA (2004) Discourse analysis and the study of organizations. *Interdisciplinary Journal for the Study of Discourse* 25/1, 37–66. DOI: 10.1515/text.2005.25.1.37.
- GRUE, Jan (2014) *Disability and discourse analysis*. London: Routledge.
- HALL, Christopher/Patrick SMITH/Rachel WICAKSONO (2011) *Mapping Applied Linguistics: A guide for students and practitioners*. New York: Routledge.
- HARRINGTON, Kate et al. (2008) *Gender and Language Research Methodologies*. New York: Palgrave Macmillan.
- HOLMES, Janet/Miriam MEYERHOFF (2003) *The handbook of language and gender*. Oxford: Blackwell.
- HOLMES, Janet/Miriam MEYERHOFF (2015) Discourse in the Workplace. D. Schiffrin/D. Tannen/H.E. Hamilton (eds.), *The handbook of Discourse Analysis*. Hoboken: John Wiley & Sons, 880–901.
- HORD, Levi (2016) Bucking the linguistic binary: gender neutral language in English, Swedish, French, and German. *Western Papers in Linguistics/Cahiers linguistiques de Western*, Vol. 3, Article 4.
- HUMMERT, Mary Lee et al. (2004) The role of age stereotypes in interpersonal communication. J.F Nussbaum/J. Harwood (eds.), *The Handbook of Communication and Aging*. Mahwah: Lawrence Erlbaum Associates, 91–114.
- JACKSON, Jane (2014) *Introducing Language and Intercultural Communication*. London: Routledge.
- JONES, Rodney (2013) Data collection and transcription in discourse analysis: a technological history. K. Hyland/B. Paltridge (eds.), *The Bloomsbury Companion to Discourse Analysis*. London: Bloomsbury Publishing, 9–21.
- KÄRREMAN, Dan (2014) Understanding organizational realities through discourse analysis: the case of discursive pragmatism. *Journal of Business Anthropology* 3/2, 201–215. DOI: 10.22439/jba.v3i2.4708.
- KOESTER, Almut (2010) *Workplace discourse*. London: Continuum.
- KRUG, Manfred/Katrin SELL (2013) Designing and conducting interviews and questionnaires. M. Krug/J. Schlüter (eds.), *Research methods in language variation and change*. Cambridge: Cambridge University Press, 69–95.
- KUBLER, Sandra/Heike ZINSMEISTER (2015) *Corpus Linguistics and Linguistically Annotated Corpora*. New York: Bloomsbury Academics.
- LEGARD, Robin/Jill KEEGAN/Kit WARD (2003) In-depth Interviews. J. Ritchie/J. Lewis (eds.), *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London: Sage, 138–169.
- LEVON, Erez (2013) Ethnography and Recording Interaction. R. Podesva/D. Sharma (eds.), *Research Methods in Linguistics*. New York: Cambridge United Press, 195–215.
- MAURANEN Anna/Elina RANTA (2009) *English as a Lingua Franca: Studies and Findings*. Newcastle: Cambridge Scholars.

- MAYR, Andrea (2008) *Language and power, an introduction to institutional discourse*. London: Continuum.
- MAYR, Andrea (2015) Institutional discourse. D. Schiffrin/D. Tannen/H.E. Hamilton (eds.), *The handbook of Discourse Analysis*. Hoboken: John Wiley & Sons, 753–774.
- MCGUIRK, Pauline M./Phillip O'NEILL (2016) Using questionnaires in qualitative human geography. I. Hay (ed.), *Qualitative Research Methods in Human Geography*. Don Mills: Oxford University Press, 246–273.
- NEUMAN, Laurence (2004) *Social Research Methods: Qualitative and Quantitative Approaches*. Harlow: Pearson Education.
- ÖZBILGIN, Mustafa/Ahu TATLI (2011) Mapping out the field of equality and diversity: rise of individualism and voluntarism. *Human Relations* 64/9, 1229–1253. DOI: 10.1177/0018726711413620.
- PILLER, Ingrid (2012) Intercultural Communication: An Overview. C. Paulston/S. Kiesling/E. Rangel (eds.), *The Handbook of Intercultural Discourse and Communication*. Malden: Wiley-Blackwell, 3–18.
- PODESVA Robert/Devayani SHARMA/Elizabeth ZSIGA (2013) Sound Recordings: acoustic and articulatory data. R. Podesva/D. Sharma (eds.), *Research Methods in Linguistics*. New York: Cambridge United Press, 169–194.
- RANTA, Elina (2009) Syntactic Features in Spoken ELF: Learner Language or Spoken Grammar? A. Mauranen/E. Ranta (eds.), *English as a Lingua Franca: Studies and Findings*. Newcastle: Cambridge Scholars, 84–106.
- RODRÍGUEZ, Cristina M. (2006) Language diversity in the workplace. *Northwestern University Law Review* 100, research paper no. 06-35.
- SARDAR, Ziauddin (2018) *The language of equality*. Manchester: Equal Opportunities Commission.
- SCHILLING, Natalie (2013) Ethics in Linguistic Research. R. Podesva/D. Sharma (eds.), *Research Methods in Linguistics*. New York: Cambridge United Press, 96–115.
- SHEN Jie et al. (2009) Managing diversity through human resource management: an international perspective and conceptual framework. *The International Journal of Human Resource Management* 20/2, 235–251. DOI: 10.1177/1038411111414984.
- TALBOT, Mary (2010) *Language and Gender*. Cambridge: Polity Press.
- TATLI, Ahu (2011) A multi-layered exploration of the diversity management field: diversity discourses. *British Journal of Management* 22, 238–253. <https://doi.org/10.1111/j.1467-8551.2010.00730.x>.
- TEN THIJE Jan D./Robert MAIER (2012) Managing cultural and linguistic diversity in multiple organisational settings: Editorial. *Journal of Multilingual and Multicultural Development* 33/7, 629–641. DOI: 10.1080/01434632.2012.713960.
- WIERSMA, Wybo (2011) *The Validity of Surveys: Online and Offline*, Oxford Internet Institute. [Accessed on 6 March 2024]. https://www.academia.edu/2373876/The_VValidity_of_Surveys_Online_and_Offline.

POVZETEK

PRIDOBIVANJE PODATKOV V PRAKSI: IZVEDLJIVOST UPORABE RAZISKOVALNIH METODOLOGIJ V DVEH JEZIKOSLOVNIH ŠTUDIJAH

Pridobivanje podatkov je temeljni vidik raziskovalnih metodologij v sociolingvistiki in jezikoslovju nasploh. Medtem ko se nekateri raziskovalci opirajo le na že dostopne podatke, se drugi odločajo za samostojno pridobivanje podatkov. Metode pridobivanja podatkov so različne – od intervjujev, vprašalnikov, pa do vodenja fokusnih skupin. Lahko jih uporabimo kot glavni vir podatkov ali le za dopolnitev obstoječega gradiva, kar največkrat izboljša kakovost raziskave in hkrati obogati raziskovalčevo osebno izkušnjo.

Namen prispevka je predstaviti izkušnji raziskovalk, ki sta v okviru svojih raziskovalnih projektov uporabili različne metodologije pridobivanja podatkov, da bi zagotovili ustreznost gradiva za analizo pisne oz. govorjene angleščine z dveh različnih specializiranih področij. Avtorici predstavita svoji raziskovalni področji in skupnosti, v katerih sta izvedli raziskavi, ter tako orišeta okoliščine svojega raziskovalnega dela. Nato opišeta postopke, ki sta jih izvedli v okviru raziskav. Njun cilj je predstaviti odločitve, ki sta jih sprejeli v času raziskav, spremembe v pristopih k pridobivanju podatkov z namenom ugotavljanja primernosti uporabljenih metodologij in rezultate raziskovalnega dela.

Avtorici želita z empiričnim uvidom v pridobivanje podatkov v različnih skupnostih prispevati k raziskavam in literaturi s področja uporabnega jezikoslovja. Njun prispevek je zasnovan kot kratek vodnik, ki raziskovalce spodbuja k lastnemu pridobivanju podatkov in jim pri tovrstnem delu tudi pomaga.

Ključne besede: raziskave s področja uporabnega jezikoslovja in sociolingvistike, korpusno jezikoslovje, pridobivanje kvalitativnih podatkov, jezikoslovne raziskovalne metode, strukturirani intervjuji in vprašalniki

ABSTRACT

HANDS-ON DATA GATHERING: VIABILITY OF RESEARCH METHODOLOGIES IN TWO LINGUISTIC STUDIES

Data collection is a fundamental aspect of linguistic and sociolinguistic research methodologies, and while many scholars may choose to gather their own data, others tend to rely solely on data that are readily available. There are various different methods of elicitation, including conducting interviews, questionnaires or leading focus groups to gather data, either as the main source of information or to complement existing material, which will most likely enhance both the quality of the research as well enriching the researcher's personal experience.

This paper aims to elucidate the experience gained by two researchers who implemented different data-collection methodologies during their research projects, to gather suitable data and conduct their analysis of the English language – one written the other oral – in two diverse, yet specialized, settings. They introduce their subjects of study and the communities in which the studies took place in order to help the reader get acquainted with the context, and then proceed to describe the processes they carried out. Their aim is to offer a retrospective on the decisions made, the evolution of their data-gathering approaches to determine the suitability of the focal methodologies, and the results.

The aim of the paper is to contribute to the literature and applied linguistics research fields by disseminating empirical information on the collection of data within different communities of practice, so that it might be used, as a brief guide, to encourage and assist fellow researchers should they decide to embark in their own data collection projects.

Keywords: applied linguistics and sociolinguistics surveys, corpus linguistics, qualitative data gathering, research methods in linguistics, structured interviews and questionnaires