



Heriot-Watt University  
Research Gateway

## Can ethnographic findings become corpus-studies data? A researcher's ethical, practical and scientific dilemmas

### Citation for published version:

Angelelli, CV 2018, 'Can ethnographic findings become corpus-studies data? A researcher's ethical, practical and scientific dilemmas', *The Interpreters' Newsletter*, no. 22, pp. 1-15.  
<https://doi.org/10.13137/2421-714X/20735>

### Digital Object Identifier (DOI):

[10.13137/2421-714X/20735](https://doi.org/10.13137/2421-714X/20735)

### Link:

[Link to publication record in Heriot-Watt Research Portal](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

The Interpreters' Newsletter

### General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Can ethnographic findings become corpus-studies data? A researcher's ethical, practical and scientific dilemmas

CLAUDIA V. ANGELELLI

Heriot-Watt University, Edinburgh campus (UK)

## Abstract

*Healthcare interpreting, performed via tele/video-conference or face-to-face interactions is complex. Research in healthcare interpreting has contributed to our understanding of this practice (Metzger 1999; Davidson 2001; Angelelli 2004, 2011, 2012; Baraldi/Gavioli 2012; Meyer 2012). Access to cross-cultural/linguistic interactions between provider/patient mediated by interpreters is essential to study intercultural/linguistic healthcare communication. Access to naturalistic data, however, is not always feasible. Therefore, researchers rely more and more on secondary data for analysis. This paper discusses ethical, practical and scientific dilemmas experienced when assessing the feasibility of turning ethnographic data into data for corpus studies. Firstly, after an introduction and a concise review of the principles underlying ethnography, the original studies are explained briefly to contextualize the data. These studies are: a) an ethnography (Spanish-English) of a medical interpreting unit and b) two case studies (Cantonese/Hmong-English) conducted in a total of three public hospitals in the United States. Secondly, a discussion on using data for a different purpose than the original one, and the resulting ethical, practical and scientific dilemmas will be presented. The goal is to reflect on and examine if the opportunities to advance science may outweigh the issues raised in this paper and if it would be ethical to proceed.*

## Keywords

Healthcare interpreting, ethnography, corpus data, ethics.

## Introduction

An interpreter-mediated healthcare encounter is a private encounter between patients (who may or may not be accompanied by family members or friends), providers, and interpreters on healthcare topics. Topics discussed range from explanations on the value of a procedure (e.g. an amniocentesis test for a pregnant woman over 35), to effects of a complex treatment (e.g. chemotherapy), as well as their cost, access and feasibility. In these discussions, the patient rather than the interpreter or the provider is the most vulnerable party of all (Zinn 2013). In contrast to public encounters in which any of the three interlocutors may be involved (e.g. a court of law where a doctor gives expert witness testimony; or a conference, in which the interpreter interprets or the patient is a participant), the interpreter-mediated healthcare encounter is a private one. This means that compared to a public setting, there are no witnesses. This affords the interlocutors a certain degree of freedom in how they may act out their roles (Angelelli 2004: 74). Being a private encounter it also means that access to observe it and collect data from it is not readily available. Procedures to apply for access and compliance with regulations for the protection of human subjects vary from country to country.

In medical private encounters, access to collect data is feasible if pertinent ethics requirements are met (e.g. obtaining informed consent from the parties; keeping data confidential, etc.). In some countries (e.g. the United States) approval processes and clearance to start collecting data may take three to four months (depending on the ethical committee meeting schedule). Thus, by the time the researcher gets approval and then determines feasibility, it is not rare to see six months or more time invested in gaining access and building trust. Ethnographies (Fetterman 2010) as well as case studies (Yin 2009) require long and sustainable efforts on the part of researchers, both for data collection and analysis. It is not unusual to read ethnographers reporting being flooded by data or drowning in data pools (Le Compte/Schensul 1999). Data collection efforts sustained through long periods of time may be at odds with study timelines or requirements (e.g. university funded scholarships for doctoral students may not accommodate longitudinal studies). And, even when practical and logistic requirements are met, there is still an element of uncertainty in data collection efforts (e.g. trust takes time to build; participants leaving the site, a sudden change in a participant's life with whom trust was established may unexpectedly render previous efforts not valuable any more, etc.). All of this makes ethnographical data even more precious, as it may be difficult to obtain a second chance to go back to the site and find the same participants with whom the ethnographer worked to be interviewed and clarify some points.

Sitting on a data gold mine, ethnographers look at ways in which they can enhance the power of their data, their significance, the different types of questions that could be asked if the data were analyzed from different perspectives, compared to other similar sites/participants, etc. and still be true to ethnographic principles. In so doing, authors of ethnographic studies may encounter trials and tribulations. This article discusses some of the issues that arise during a journey

that started conducting ethnography and ended in preparing/adapting data to be used as a corpus. Like with any change in processes, there are losses and gains in this journey. My goal in writing this article is to contribute to an on-going discussion (Angermyer *et al.* 2012; Bendazzoli 2012; House *et al.* 2012) with researchers who face similar dilemmas and possibilities, as well as to contribute to the body of knowledge created by researchers already working with corpus data on community interpreting (see, for example, <http://www.yorku.ca/comindat/comindat.htm> or <http://pagines.uab.cat/tipp/>) for different purposes, such as teaching of both interpreting (Meyer 1998; Bendazzoli *et al.* 2011; Bührig *et al.* 2012) and translation (Munday 1998; Oloham 2004; Rabadán *et al.* 2009; House 2011).

To ground my contribution, I will first present an overview of ethnography (§ 2) followed by a brief contextualization of the two studies under consideration (§ 3). The remaining part of this paper discusses a journey of reflection. Adapting ethnographic data and getting them ready to build a corpus, as well as the resulting challenges and opportunities encountered along the way, triggered this reflection piece.

## 1. Ethnography as a research method

The term ethnography, initially used to name the work performed by anthropologists, travelled from Anthropology to Sociology and other Social Sciences (communication, education) and has become more and more discussed in Translation and Interpreting Studies although at times, it is discussed partially as equated with qualitative methods, e.g. ethnographic methods or ethnographic interviews (Koskinen 2006; Angelelli 2015) or confused with case studies (Hale 2007: 63). The Merriam Webster dictionary defines ethnography as “the study and systematic recording of human culture” to also include the resulting descriptive work produced (Merriam-Webster 2017). The term ‘ethnography of communication’ (Hymes 1964, 1974) is worth discussing as it provides a frame of reference for studying language as used by people, whether at the level of society or an organization. Therefore, Hymes’ work offers an important theoretical and analytical lens to learn about ways of speaking of different speech communities (physicians, patients, translators, interpreters), as well as about the participants, or channels used etc. Ethnography and Hymes’ framework have been applied in both Translation and Interpreting Studies to study translators (Asare 2015) and interpreters’ work (Angelelli 2000; Mack 2002) and to compare communicative events (monolingual and interpreted ones) in, for example, an educational setting (Valdés *et al.* 2000) and a medical setting (Angelelli 2004: 34-40).

Conducting an ethnography affords the researcher an emic (insider) rather than an etic (outsider) perspective on the data (Morris *et al.* 1999: 783). As time goes by the ethnographer’s view shifts from the one of an outsider, or the observer to the one of the local, the native, the member of the community observed. By gaining a similar perspective to that of an insider the researcher is in a better position to learn, interpret and even question the ways of doing (e.g. speaking, behaving) of the members of the community observed. This is accomplished by a

focussed, rigorous, sustainable and continuous effort which is necessary to learn about the ways of doing of *the other* observed. The ethnographer makes sense of patterns of behavior and learns to distinguish between typical and atypical ones. Ethnographers walk into a community (e.g. linguistic, occupational), to systematically and constantly observe, learn and record these ways of doing, of speaking, etc. In so doing, ethnographers take a naturalistic approach to data and do not manipulate them. This means that when ethnographers enter a site, they may not have a definite research question in mind for which they want an answer; they do not enter a site to collect data in order to accept or reject a hypothesis. Instead, they take in all information and knowledge, even if at times it does not make sense to them and it was not expected to learn from it and understand it. Their analytical work is led by the data (knowledge and information) gathered, not the other way around. It is important to bear this distinction in mind. It is not unusual to confuse an ethnography with other qualitative types of research such as case studies (especially longitudinal ones). To understand the difference means to understand the intimate relationship between the researcher and the reality observed/studied. The researcher enters the site with a hunch or a curiosity rather than with a definite goal (answering a specific research question or conducting an experiment). This implies not only the use of a specific research approach, method, timeline, analytical lens or paradigm. It also implies a different way of conceptualizing and organizing data.

While doing ethnography, collecting data and analyzing findings are iterative processes. Ethnographers' reports include data gathered through extended field observations (both participant and non-participant ones), and ethnographers are said to be the most important research tool in the study. Analytical categories are not imposed on data but rather emerge from it based on frequency and typicality. Given all of these, one can understand why the question of turning ethnographic data into a corpus merits some serious considerations, as all concepts need to travel across paradigms and research cultures.

Traditional constructs of objectivity and detachment, which have been central to a positivist research paradigm may be constructed differently in other paradigms (e.g. post-positivists) while analyzing the same data set. Ethnographic data on communication result from a specific discourse community immersed in a specific context, meant to be studied from an emic perspective within its context. When turned into a corpus many of the fundamental notions of the conceptual framework that guided the ethnography of communication, for example, and using Hymes' terms (1974: 45-62), the scene, the setting, the participants, the purposes, or the channels of the communicative event studied may no longer be in the corpus. This may occur, of course, only if the corpus is limited to the transcripts and audio files as the new researcher following the path of the ethnographer is facing the message content (Hymes 1974: 55) only.

While extralinguistic information could be included in the form of a header in the transcript, the range and scope of this information are often limited in order to make this kind of annotation more user friendly. And issues like this (limitations imposed by the software or the community of users), are precisely the ones with which an ethnographer has to grapple when thinking of turning

ethnographic data into a corpus. Are we really talking about ethnographic data? If we limit the data to accommodate it to the new tool used to analyze it, are we then not changing the scope and nature of the data? However, these fundamental notions would be available for the new researcher if the corpus were to include all artifacts, interviews, pictures, etc. which is something ethnographers do when using qualitative software such as NVivo (<http://www.qsrinternational.com>) or The Ethnograph (<http://www.qualisresearch.com>) to analyze and organize qualitative data in an electronic format. Qualitative researchers have analyzed and compared specific software that can help the analyst in the work at hand (Gilbert 2002; Richards/Richards 1991; Woods *et al.* 2015).

The eyes and lens of other researchers examining the corpus differ from the eyes of the original ethnographer as they have not had continuity on site. From an ethnographic view, removing sustained presence at the site and conducting check-visits for a specific purpose instead (e.g. querying the data for discourse fillers), may not make as much sense as from a corpus linguistic one. So, in the end, the data transfer may indeed constitute a journey across paradigms and research cultures rather than a data crisis or a turning point. Given that we are discussing interpretation of linguistic and sociolinguist data across languages and cultures, a reference to a term in one language and its journey towards another language is worth mentioning here: when looking up the term *crisis* in Chinese, two characters are used to depict it. These two characters represent challenge and opportunity. In this article, after giving the reader a brief description of the ethnographic data, I turn to the challenges and opportunities of using ethnographic data to build a corpus.

## 2. Brief contextualization of the original studies

The data which will form the bulk of the (forthcoming) California Hospital Interpreting Corpus (CHIC) was collected by the author in three public hospitals in California for two separate original studies with distinct purposes.

### 2.1 Original study 1

The data for Spanish-English interpreted communicative events results from an ethnographic study of Spanish-English interpreted medical communication conducted in a public hospital (California Hope, see Angelelli 2004) around the Bay area, in California, between 1998 and 2000. The purpose of the ethnography of communication (Hymes 1974) was to learn the ways of speaking of linguistically and culturally diverse discourse communities when discussing private health matters mediated by interpreters. Communication between providers and patients was brokered by staff medical interpreters.

To accomplish the purpose of the original study the researcher used multiple methods of data collection and multiple analysis. These are: site observations, observation of interpreted communicative events (ICEs, Angelelli 2000), ethno-

graphic interviews with participants, data interpretation interviews, conceptual memos, notes, artifacts and questionnaires. The subsequent transcription and translation of the 392 Spanish-English interpreted communicative events yielded 2,500 electronic pages of data (4.5 MB), for a total of 521,717 tokens in 153,200 lines. In addition to the interpreted-communicative events, the researcher conducted interviews with each of the participating interpreters and their manager on issues related to their work and role. Those transcripts are available in English to be added to the corpus.

## 2.2 Original study 2

The data for Hmong and Cantonese interpreted communicative events were collected during two case studies conducted in two public hospitals in the Central Valley area of California. These studies were part of a larger project funded to develop a battery of tests to evaluate medical interpreters in California in a meaningful, valid and reliable way (see Angelelli 2007). In the year 2000, with funding and support from The California Endowment, the Connecting World Partnership (CWP) a consortium of five organizations in California, commissioned the author to develop a Language-Proficiency (LP) test and an Interpreter Readiness (IR) test in three languages: Spanish, Hmong, and Cantonese. The members of CHC were Asian Health Services; Healthy House with a MATCH coalition; Las Clínicas de Salud del Pueblo; PALS for Health and Vista Community Clinic. The hospitals in the Central Valley area were chosen because of the number of Hmong and Cantonese patients that visit the hospital. The case studies were conducted in Spring semester of 2001 by a team of three: principal investigator, research assistant and interpreter. During that time 60 ICEs were collected for Cantonese and 56 for Hmong. Health providers, patients and interpreters communicated over health issues. Interactions were recorded, observed and partially transcribed and translated to meet the requirements of the original project. The transcriptions and translations need to be completed and transferred into a corpus format. This can be accomplished once funding is secured.

In the next section I discuss some of the considerations, challenges and opportunities faced during the journey.

## 3. From ethnographic data to a corpus: challenges and opportunities

Before starting the process of considering the original data for an electronic corpus to be shared, the researcher had asked specific questions of the ethnographic data. These questions were posed after the ethnographic study was finalized. Questions were both at the macro and micro level. At the macro level, for example, questions related to ways in which interpreters construct understanding (or misunderstanding) among patients and providers while speaking about delicate issues, such as terminating a pregnancy (Angelelli/Geist-Martin 2005) or using a pain-rating scale (Angelelli 2012). At the micro level, questions involved dis-

course bundles (Biber/Conrad 1999), set expressions and collocations such as the use of the term *chronic illnesses* (Angelelli 2011) or the use of a pain scale (Angelelli 2012). It was feasible for the researcher to ask specific questions from a Word database and to conduct searches together with a co-author, simply because the ethnographer was very familiar with the data. In addition, the time lag between the end of the ethnography and the writing was relatively short. As I had not experienced working through my own database with the help of other resources such as annotations and indexes produced by others, I cannot evaluate my experience for those searches in terms of convenience, speed, etc. I can see, however, some evolution of my thinking between now and then.

### 3.1 Scientific dilemmas

#### 3.1.1 Scope

Transferring ethnographic data into a corpus raises some questions about scope. Ethnographic studies are known for producing large amounts of data. Decisions have to be made as to the data that can be transferred. If only transcripts are included in the corpus, and the rest of the data sources that helped the ethnographer perform a thick interpretation of the data are not, then this calls into question the scope of subsequent studies based on the data. If only transcripts are transferred, then interpretation of the corpus analyst and the ethnographer could not be comparable. If all of the ethnographic data (including observations and recordings, pictures, artifacts, data interpretation interviews, ethnographic interviews, as well as transcripts of the interpreted communicative events, conceptual memos to self) were transferred, then, the corpus analyst could be in a position to almost replicate the nature of the original study, as the corpus analyst would be getting almost the same amount and type of data as the ethnographer but would be one degree removed.

This remote and one-degree removed position of the researcher allows for new possibilities, as the corpus analyst could query the data and triangulate almost in the same way the ethnographer did but, this time it would be using secondary data and with no access to participants. This is an interesting proposition that requires further exploration, especially in relationship to the issues discussed on gaining entry to sites or building trust with participants. This statement, however, by no means suggests that accessing only the electronic data could afford the insight gained by sustained effort and time spent in the field site.

#### 3.1.2 Data sharing

While ethnography of communication has generally been a one-researcher endeavor, the ability to now digitalize data and make data available to other researchers forever changes the way in which we conceptualize this type of studies and designs. Ethnographic data now can be shared with others. This sharing has



advantages and disadvantages. The advantage is that sharing data enables multiple analyses and interpretations from different perspectives, spaces and times. It also allows for comparisons across different linguistic combinations. While the ethnographer generally validates findings with participants and, at times, it is also possible to consult with other researchers to get another perspective, an electronic database increases the opportunities to perform both these tasks beyond limit. This is an advantage that requires compromises.

The disadvantage is we are no longer dealing with ethnography and this needs to be acknowledged. Sharing ethnographic data changes the nature of the analysis. We can no longer access the site, the context, the cues, etc. Instead, we access what can be captured in, and becomes available to us from a database. We access a product, not the iterative process. Reality becomes in some ways mediated by a dataset as well as regulated by it. It is no longer a contextualized direct observation that raises questions or produces data, such as a transcript. Now, it is a transcript (even if decontextualized) that leads the researcher. The tension between using transcripts within or without a context has already been addressed in debates between conversational and discourse analysts (for a brief overview see, for example, Wooffitt 2005; Antaki 2008). Now the transcript becomes the object of study. For some researchers this compromise is a problem, for others an opportunity. There would simply be no opportunity to discuss the data from multiple perspectives, contexts and cultural viewpoints if it were not accessible electronically. And this would be a loss.

### 3.1.3 Categories, definition of tags/annotations

Categories and patterns of data of the original study may or may not transfer directly into corpus tags and annotations. Transferring transcripts initially conceived to be used for one purpose (e.g. understand communication or to be used in a test script) to another may require some adjustments. The pros and cons of using different types of annotations have been discussed extensively in corpus linguistics and, specifically for community interpreting, more recently by Angermeyer *et al.* (2012). Therefore, instead of engaging in description of technical issues that may or may not have a solution viable for all of those who contribute to a corpus, I would like to take a more philosophical/conceptual approach in the discussion of categories, definitions or annotations.

Deciding on categories and annotations *a priori*, or seeing those of others before diving into the data may constitute a philosophical dilemma for an ethnographer trying to access an existing corpus or preparing his/her own. In ethnography, categories emerge from data and are not a consideration *a priori*. This does not mean however that we, as a community of researchers with shared interests, could not access each other's data with different purposes and using different lenses. Ethnographers can share their categories and make use (or not) of pre-existing categories applied by other researchers to the original ethnographic corpus. Corpus analysts may find exciting opportunities in accessing ethnographic data and transfer pre-existing categories to it and obtain results. Discourse ana-

lysts may transfer pre-existing categories from a study (e.g. discourse bundles in medical communication) with the same language combination, and query the ethnographic data for those occurrences. And, even when ethnographers may not necessarily benefit from looking at other corpora, as that may not be their approach to research, and when corpus annotations depend on the aim of one's study, making ethnographic data available to other types of researcher will undoubtedly contribute to science even if it poses a dilemma.

#### 3.1.4 Dataset

Turning ethnographic data into a searchable dataset offers several advantages to the researcher. Firstly, by turning notes, pictures and artifacts into a searchable electronic corpus the researcher is able to access data faster and easily and can see issues/ask questions that perhaps would have remained unseen/unasked. Secondly, having the ability to compare and contrast naturalistic data (e.g. recordings and resulting transcriptions) with researcher intake and interpretation of such data in the form of conceptual memos, allows for the ethnographer to conduct some type of 'intra-rater reliability check'. Although it is understood that notions of objectivity differ significantly across research paradigms, having the chance to double check and validate one's own perceptions, recalls and insight against a wide range of data sources in a second, makes an invaluable difference that should not be taken lightly. Thirdly, having all data sources in a searchable database allows for quicker verification of emerging categories that can later be used to organize the data for the final report/story. The more we use corpora to analyze interpreted interactions the more we learn about them. Avoiding a top-down approach to categories is possible with researchers' awareness and self-monitoring. So this may not constitute a dilemma. The issue of time and cost of turning ethnographic data into an electronic data set remains.

#### 3.2 Ethical dilemmas

To obtain permission to conduct a study, gain entry to a site, and collect data, whether in the form of field notes and observations or recording of interactions, researchers file ethics and human subject protocols with the Internal Review Board of the university that hosts the researcher and the healthcare organization where the study takes place. The protocol includes a clear explanation of the goal and objectives of the study, the duration, the selection of participants, the materials and procedures to be used, how researchers plan to explain the study to participants, etc. Also in these protocols filed with the Internal Review Board for the Protection of Human Subjects, researchers have to explain how they plan to store, make use and dispose of their data. Different rules apply in different countries as to the protection of human subjects and the confidentiality of the data. In the United States, for example, participants have to give informed consent as they sign and date a specific form that explains the goal and nature of the

study, the duration, the type of data that will be collected, how it will be used, etc. and who is the party responsible for the data collection, use and storage. In these forms participants consent to different things (e.g. to be observed, audio or video recorded [or both], interviewed, etc.) separately. Because the forms may not include a separate section on making data available to other researchers or sharing them in a database or corpus, many times researchers face the dilemma of making decisions based on their own intuitions (e.g. if stripping data from all personal identifiers takes care of confidentiality for research purposes, can this also be applicable to including data in a corpus, even if there cannot be any control on the purposes for which data is used?). Evidently as corpora become available and data is shared among researchers, various constructions of ethics and protection of human subjects interact. The need to have a shared conceptualization of ethics and protection of human subjects around the world is essential for shared scientific projects.

### 3.3 Practical dilemmas

#### 3.3.1 Sensitive settings: better access via corpus

The data in the CHIC comes from private provider/patient interpreted encounters. The healthcare setting is a site in which sensitive conversations may take place. Discussions of infectious diseases, life and death, amputations, terminal illnesses, complicated treatment and occupational therapy for paraplegic patients are not only sensitive, but, at times, they can be humiliating for the patients. These discussions do not always lend themselves to be observed and recorded. They require some degree of trust between the researcher and the patient on the one hand (as the patient must be willing to share confidential information), as well as trust on the part of all interlocutors as to the use that the researcher will make of the data. In addition, sensitive topics and interactions require that the researcher observing overcomes emotions and be capable to detach the self from the issue and to focus on the analysis of the object of study (which generally is the communication about the issue at hand, rather than the issue itself). Distance and detachment from the scene and participants are sometimes helpful to focus attention mostly on the communicative issue at hand. Many times the days in the field are difficult and stressful. After so much time spent with participants, detachment from issues and feelings becomes harder. This is an area in which the advantages of working from a corpus rather than from direct observations could be stressed.

#### 3.3.2 Time lag and data access

An electronic format and an organized database will make access more feasible than going back to original data and searching manually. Revisiting the data in machine-readable format after an amount of time has passed may have advan-

tages also for the ethnographer, not only for other researchers. If the time lag in between the creation of a corpus and the original data collection is considerable, the ethnographer may not be able to rely on memory any more and, in addition, the ethnographer may be able to trace changes in her own perspectives or interpretation of original data. This would add a new dimension to intra-rater reliability of data as it could track re-interpretations and potential changes or lack thereof after specific periods of time.

#### 4. Conclusion and implications

The passing of time sometimes allows for learning and change. At the time of *doing ethnography* at California Hope, building a corpus of authentic interactions was not a possibility. At that time, the thought of being true to ethnographic tradition, as well as the responsibility to comply with specific rules and regulations on human subject protection, data gathering, storing and sharing guided my work, my analysis and my reporting. Later on, the same thought and the same rules and regulations prevented me from sharing data. Conceiving the idea that, in due course, the California Hope database could become something different than ethnographic data, that it could feed into a corpus, was attractive and challenging. Time was necessary to produce clear processes and policies as well as to honor previous agreements of confidentiality. Once the time to comply with specific data storing and management has elapsed and previous agreements have been honoured, we can entertain other possibilities and engage in scientific conversations about different kinds of analyses and sets of data. The thought that an original ethnographic database can become a corpus accessible to others affords us more opportunities and may, most certainly, contribute to advancing our knowledge.

In this article we discussed the journey faced by a researcher while considering turning ethnographic data into a corpus. The hope is to have contributed to a conversation on the possibilities and challenges of turning ethnographic data into corpus data. Since access, time and cost are often discussed as obstacles to having databases of interpreted communicative events readily available, sharing data in the form of corpora seems to be a viable option for researchers. Sharing data and making it available to others, however, also implies agreeing on a series of rules for all the steps of the process (e.g. from transcription conventions and data protections to annotations) across languages and cultures. All of these issues have been discussed in the literature in Translation and Interpreting Studies, from the first call made by Mona Baker (1993) to the latest publications in both community and conference interpreting (Angermeyer *et al.* 2012; Bendazzoli *et al.* forthcoming; Russo *et al.* 2018; Straniero Sergio/Falbo 2012). Much progress has been made. While researchers think more and more along the lines of making data available in the form of corpora, four issues remain constant: complying with national/international laws, rules and regulations governing protection of human subjects and data storage/sharing, achieving a shared understanding of ethics (specifically across diverse communities), time and cost. As discussed

above, one should not underestimate the amount of time and cost involved in preparing data for a corpus. And, most importantly, this huge task can hardly be an add-on or an a-posteriori thought of any project.

This discussion has implications for students and researchers of Interpreting Studies, as well as for funding agencies and research sites. Students and researchers who may be studying phenomena embedded in interpreted interactions will benefit enormously from existing corpora both for obtaining data or contributing their own. Therefore, anticipating effort, time and cost for building a corpus and factoring them in their on-going projects may help students and early-career researchers take care of corpus building at the same time as they conduct their studies. Funding agencies should continue assigning (and even increase) funding for digital humanities. Transforming data into a corpus is time consuming and universities generally do not give credit for such an enduring task which is not considered primary research. And without funding to cover expenses or time, the technical/practical part of building a corpus may still not be feasible.

## References

- Angelelli C.V. (2000) "Interpreting as a communicative event: a look through Hymes' lenses", *Meta* 45/4, 580-592.
- Angelelli C.V. (2004) *Medical Interpreting and Cross-cultural Communication*, London, Cambridge University Press.
- Angelelli C.V. (2007) "Accommodating the need for medical interpreters: The California Endowment Interpreter Testing Project" (This work was made possible by The California Endowment Grant G00005318), *The Translator* 13/1, 63-82.
- Angelelli C.V. (2011) "Can you ask her about chronic illnesses, diabetes and all that?", in C. Alvstad / A. Hild / E. Tiselius (eds) *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies*, Amsterdam/Philadelphia, John Benjamins, 231-246.
- Angelelli C.V. (2012) "Challenges in interpreters' coordination of the construction of pain", in C. Baraldi / L. Gavioli (eds) *Coordinating Participation in Dialogue Interpreting*, Amsterdam/Philadelphia, John Benjamins, 244-260.
- Angelelli C.V. (2015) "Ethnographic methods", in F. Pöchhacker (ed.) *The Routledge Encyclopedia of Interpreting Studies*, New York, Routledge.
- Angelelli C.V. / Geist-Martin P. (2005) "Enhancing culturally competent health communication: constructing understanding between providers and culturally diverse patients" in E. Berlin Ray (ed.) *Health Communication in Practice. A Case Study Approach*, New Jersey, Lawrence Erlbaum Associates, 271-284.
- Angermeyer P. / Meyer B. / Schmidt T. (2012) "Sharing community interpreting corpora: a pilot study", in T. Schmidt / K. Worner (eds) *Multilingual Corpora and Multilingual Corpus Analysis*, Amsterdam/Philadelphia, John Benjamins, 275-294.

- Antaki C. (2008) "Discourse analysis and conversation analysis", in P. Alasuutari / L. Bickman / J. Brannan (eds) *The SAGE Handbook of Social Research Methods*, London, Sage, 431-446.
- Asare E. (2015) "Ethnography of communication", in C.V. Angelelli / B.J. Baer (eds) *Researching Translation and Interpreting*, New York, Routledge, 212-219.
- Baker M. (1993) "Corpus linguistics and translation studies – implications and applications", in M. Baker / G. Francis / E. Tognini-Bonelli (eds) *Text and Technology: In Honour of John Sinclair*, Amsterdam/Philadelphia, John Benjamins, 233-250.
- Baraldi C. / Gavioli L. (2012) "Introduction: understanding coordination in interpreter-mediated interaction", in C. Baraldi / L. Gavioli (eds) *Coordinating Participation in Dialogue Interpreting*, Amsterdam/Philadelphia, John Benjamins, 1-22.
- Bendazzoli C. (2012) "From international conferences to machine-readable corpora and back: an ethnographic approach to simultaneous interpreter-mediated communicative events", in F. Straneiro Sergio / C. Falbo (eds) *Breaking Ground in Corpus-based Interpreting Studies*, Bern, Peter Lang, 91-118.
- Bendazzoli C. / Russo M. / Defrancq B. (eds) (forthcoming) *Corpus-based Interpreting Studies: The State-of-the-art*. Special issue of *inTRALinea* 20.
- Bendazzoli C. / Sandrelli A. / Russo M. (2011) "Disfluencies in simultaneous interpreting: a corpus-based analysis", in A. Kruger / K. Wallmach / J. Munday (eds) *Corpus-based Translation Studies: Research and Applications*, London, Continuum: 282-306.
- Biber D. / Conrad S. (1999) "Lexical bundles in conversation and academic prose", in H. Hasselgard / S. Oksefjell (eds) *Out of Corpora: Studies in Honour of Stig Johansson*, Amsterdam, Rodopi, 181-189.
- Bührig K. / Kliche O. / Meyer B. / Pawlack B. (2012) "The corpus 'Interpreting in Hospitals': possible applications for research and communication training", in T. Schmidt / K. Worner (eds) *Multilingual Corpora and Multilingual Corpus Analysis*, Amsterdam/Philadelphia, John Benjamins, 395-318.
- Davidson B. (2001) "Questions in cross-linguistic medical encounters: the role of the hospital interpreter", *Anthropological Quarterly* 74/4, 170-178.
- Fetterman D. (2010) *Ethnography: Step-by-Step*. 3<sup>rd</sup> edition, Thousand Oaks CA, Sage.
- Gilbert L.S. (2002) "Going the distance: 'closeness' in qualitative data analysis software", *International Journal of Social Research Methodology* 5, 215-228.
- Hale S. (2007) *Community Interpreting*, Basingstoke, Palgrave Macmillan.
- House J. (2011) "Using translation and parallel text corpora to investigate the influence of global English on textual norms in other languages", in A. Kruger / K. Wallmach / J. Munday (eds) *Corpus-based Translation Studies: Research and Applications*, London, Continuum, 187-208.
- House J. / Meyer B. / Schmidt T. (2012) "CoSi - a corpus of consecutive and simultaneous interpreting", in T. Schmidt / K. Worner (eds) *Multilingual Corpora and Multilingual Corpus Analysis*, Amsterdam/Philadelphia, John Benjamins, 295-304.

- Hymes D. (1964) "Introduction: toward ethnographies of communication", in J.J. Gumperz / D. Hymes (eds) *The Ethnography of Communication*, Special issue of *American Anthropologist* 66/6, part 2, 1-34.
- Hymes D. (1974) *Foundations in Sociolinguistics*, Cambridge, Cambridge University Press.
- Koskinen K. (2006) "Going into the field: ethnographic methods in Translation Studies", in M. Wolf (ed.) *Übersetzen – Translating – Traduire*, London, Transaction Publishers, 109-117.
- Le Compte M.D. / Schensul J.L. (1999) *Analyzing and Interpreting Ethnographic Data*, vol. 5, Walnut Creek, AltaMira.
- Mack G. (2002) "New perspectives and challenges for interpretation. The example of television", in G. Garzone / M. Viezzi (eds) *Interpreting in the 21st Century: Challenges and Opportunities. Selected Papers from the 1<sup>st</sup> Forlì Conference on Interpreting Studies, 9-11 November 2000*, Amsterdam/Philadelphia, John Benjamins, 203-213.
- Merriam Webster Dictionary, <<https://www.merriam-webster.com/dictionary/ethnography>> (last accessed August 29th, 2017).
- Metzger M. (1999) *Sign Language Interpreting: Deconstructing the Myth of Neutrality*. Washington DC, Gallaudet University Press.
- Meyer B. (1998) "What transcriptions of authentic discourse can reveal about interpreting", *Interpreting* 3/1, 65-83.
- Meyer B. (2012) "Ad hoc interpreting for partially language-proficient patients: participation in multilingual constellations", in C. Baraldi / L. Gavioli (eds) *Coordinating Participation in Dialogue Interpreting*, Amsterdam/Philadelphia, John Benjamins, 99-113.
- Morris M. / Leung K. / Ames D. / Lickel B. (1999) "Views from inside and outside: integrating emic and etic insights about culture and justice judgment", *Academy of Management Review* 24/4, 781-796.
- Munday J. (1998) "A computer-assisted approach to the analysis of translation shifts", *Meta* 43/4, 542-556.
- Olohan M. (2004) *Introducing Corpora in Translation Studies*, London, Routledge.
- Rabadán R. / Labrador B. / Ramón N. (2009) "Corpus-based contrastive analysis and translation universals: a tool for translation quality assessment English → Spanish", *Babel* 55/4, 303-328.
- Richards T. / Richards L. (1991) "The NUDIST qualitative data analysis system", *Qualitative Sociology* 14/4, 307-324.
- Russo M. / Bendazzoli C. / Defrancq B. (eds) (2018) *Making Way in Corpus-based Interpreting Studies*, Singapore, Springer.
- Shlesinger M. (1998) "Corpus-based Interpreting Studies as an offshoot of corpus-based Translation Studies", *Meta* 43/4, 486-493.
- Straniero Sergio F. / Falbo C. (eds) (2012) *Breaking Ground in Corpus-based Interpreting Studies*, Bern, Peter Lang.
- Valdés G. / Chavez C. / Angelelli C.V. (2000) "Bilingualism from another perspective: the case of young interpreters from immigrant communities", in A. Roca (ed.) *Research on Spanish in the United States. Linguistic Issues and Challenges*, Somerville Mass, Cascadilla Press, 42-81.

- Woods M. / Paulus T. / Atkins D.D. / Macklin R. (2015) "Advancing qualitative research using qualitative data analysis software (QDAS)? Reviewing potential versus practice in published studies using ATLAS.ti and NVivo, 1994–2013", *Social Science Computer Review* 34/5, 597-617.
- Wooffitt R. (2005) *Conversation Analysis and Discourse Analysis*, New York, Sage.
- Yin R.K. (2009) *Case Study Research: Design and Methods*, 4<sup>th</sup> ed., Thousand Oaks, Sage.
- Zinn J. (2013) "Surgical patients: a vulnerable population", *ARON Journal. The Official Journal of Perioperative Medicine* 98/6, 647-652.