Designing studies with naturalistic tasks

DESIGNING STUDIES WITH NATURALISTIC TASKS

Christopher D. Mellinger, Nicoletta Spinolo,
Maureen Ehrensberger-Dow & Sharon O'Brien

University of North Carolina at Charlotte - University of Bologna - Zurich University of Applied Sciences - Dublin City University

This chapter presents an overview of what constitutes a naturalistic task in cognitive translation and interpreting studies (CTIS), taking into account research settings, methodologies, data collection methods, and the participants involved. Naturalistic tasks are situated along a continuum that can, at one extreme, replicate specific working conditions or tasks encountered in authentic settings, and at the other, focus on specific task elements under tightly-controlled conditions. We explore possible applications of naturalistic tasks in CTIS study designs and discuss the rationale for using naturalistic tasks instead of or in addition to other types of tasks. We reflect especially on ethical concerns and conclude with a consideration of the emerging challenges in doing research with a commitment to naturalistic tasks and settings.

When reference is made to naturalistic tasks in other disciplines interested in cognition, such as bilingualism, psycholinguistics, and neuroscience, these tasks are often described in contrast to 'abstract,' 'clinical,' or 'controlled' tasks (e.g., Chou et al., 2021; Crum, 2021; Robertson & Schmitter-Edgecombe, 2017). The term naturalistic has been used in various Cognitive Translation & Interpreting Studies (CTIS) publications to suggest a task that is closer to the lived reality of anyone engaged in interpreting and translation (e.g., Albl-Mikasa et al., 2020). Therefore, it might be considered the gold standard for testing assumptions and claims based on less-than-natural tasks in less-than-natural settings.

One of the strongest arguments for designing studies with naturalistic tasks is to increase the ecological validity of a particular research endeavor. Ecological validity, which has been defined as "the degree to which test performance predicts behaviors in real-world settings" (Salkind, 2010, p.399), is still a controversial concept, but there seems to be an agreement that "a deeper understanding of the concept can be achieved by analyzing its three dimensions: test environment, stimuli under examination, and behavioral response" (Salkind, 2010, p.399). In fact, a 'natural' task would probably not even be labeled as such because this type of task would simply be regular work activity in the usual setting where it is performed. Such activity might be studied, for example, through ethnographic research approaches and workplace research which has, among other reasons, been carried out to analyze aspects of the translation process (for an overview, see Ehrensberger-Dow & Massey, 2019). However, truly natural tasks are nearly impossible to achieve in a research environment: We know that the very act of observing an event can alter it (i.e., observer effects, discussed further in Section 4). A

'naturalistic' task or stimulus, then, can be considered to be somewhere on a continuum between artificial or contrived (i.e., verging on the unnatural) and natural.

At the other end of the continuum, certain research questions and data collection methods require highly controlled conditions (e.g., eye tracking) and experimental stimuli that might bear little resemblance to the type of work done at the workplace. Sometimes, though, the aims of a study may demand only a minimum level of intervention, such as ensuring that texts or speeches are of a similar length or display a certain number of features of particular interest for the research question. In such cases, the advantages of carrying out data collection with a naturalistic task in the controlled setting of a quiet lab can outweigh the disadvantages of potential distractions associated with observations in the field (cf. Salkind, 2010, p.399).

One point on which most CTIS researchers would agree is that naturalistic tasks must involve texts or communicative interactions that are 'non-random' (cf. Shreve, 2018) and meet most if not all standards of textuality. These include textcentered notions such as cohesion and coherence as well as user-centered notions concerning intentionality, acceptability, informativity, situationality, and intertextuality (cf. de Beaugrande & Dressler, 1981). For example, if the related activity in professional settings is associated with the translation of single words or out-of-context phrases (e.g., in localization), then translating a list of words or phrases could be a naturalistic task if the latter are non-random and can be seen to have intentionality. However, translating a list of words designed for their orthographic, morphological, or frequency properties would be a contrived task. In the case of participants chosen because they are professional translators, subtitlers, post-editors, dubbers, interpreters and so on, the stimuli should be as similar as possible to the linguistic input they normally work with.

Indeed, task difficulty in CTIS has been strongly associated with the source text (e.g., Liu & Chiu, 2009; Shreve, 2006), and an example of the challenges associated with naturalistic tasks relates to the choice of appropriate text. In the past, to reconcile differences in the background knowledge and experience that participants bring to an empirical study, translation researchers have often resorted to texts from the field of journalism and interpreting researchers to pre-recorded speeches that are assumed to be quite generic and familiar to almost everyone, if not in content, at least in style. The reality, however, is that such texts and speeches are rarely translated or interpreted in full to produce parallel versions in several languages. Instead, they are adapted, recontextualized, and reframed to accommodate the needs of the target audience in a particular context. For nonspecialists (i.e., people working outside the domain of journalism and media), producing an adequate rendering in another language would be an unrealistic task. Defining naturalistic tasks and finding source texts or speeches that would be appropriate for the participants' background and provide them with an opportunity to demonstrate their expertise means doing preliminary research into their professional realities (i.e., workplace research; see Risku et al., 2020).

Although the situated activity of professional translation has mostly moved far away from pen, paper, and print dictionaries, the types of tasks used to explore its cognitive dimensions have not always kept pace. For example, most professional translators in wealthy countries work in highly technologized settings with translation tools ranging from term bases and computer-aided translation (CAT) systems to voice-activated text editing and adaptive machine

translation. Even most students cannot imagine doing translation work without having access to the Internet to allow them to search for background materials and resolve lexical choices or terminology difficulties. A task that requires such individuals to translate using only pen and paper or just a text editor without access to any other digital resources or to interpret without preparation would be in most cases far removed from their (cognitive) life world and might seem to have little to do with their language work. Any indicators of cognitive effort, stress, emotional response, and so on could thus reflect unfamiliarity with the task rather than variables such as difficulty of the source text, time pressure, or particular translation challenges.

Conference and community interpreting have been described as examples of highly situated expert bilingual activities. As such, the demands of dealing with presentations by unfamiliar speakers or signers within the confines of the booth or dialogues occurring in specific settings can only be partly simulated by naturalistic tasks in training and research settings. As soon as the situation is not live, for example when the speakers, signers, or audience have to be imagined rather than experienced, then it is questionable how naturalistic interpreting in simulated settings can be, at least to professional interpreters. It might be easier to set up naturalistic tasks in remote interpreting scenarios, but then it is questionable how generalizable they would be to interpreting that is performed in person. The same holds for the interpreter's cognitive and affective demands, which might have a very different quality when working in simulated settings as opposed to professional contexts.

Rather than futilely attempting to define what naturalistic tasks are and how to standardize them, this chapter explores the aspects that need to be taken into account in order to do justice to claiming that a naturalistic task has been used in a particular study. These start with the hardware and software that the participant will use: an unfamiliar keyboard, console, interface, operating system, search engine, and CAT tool can all affect translation or interpreting performance and confound the results of an otherwise well-designed study. Ideally, participants use their own equipment to complete tasks that resemble their usual work (i.e., if usual behavior is of interest). Any recording equipment intended to capture everything that happens during the tasks should be as unobtrusive as the researcher is or, at least, as similar to that done in authentic settings as possible, since naturalistic tasks should not feel artificial, decontextualized, or unduly constrained.

Most researchers who engage in workplace research, including the authors of this chapter, do not refer to naturalistic tasks because they usually observe T&I professionals doing their normal work in their normal environment (e.g., Ehrensberger-Dow & Hunziker Heeb, 2016; Kuznik & Verd, 2010; Risku et al., 2020; Teixeira & O'Brien, 2017). One could argue that a good example of a naturalistic task would be to ask study participants to translate or interpret at their usual workplaces with their usual resources and tools a source text similar to the texts they usually work with (e.g., Teixeira, 2014). In order to allow the researchers to make more direct comparisons, the same source text should be used, and the same brief and background materials provided. Although it would not be considered very naturalistic for several translators or interpreters to process the same text under normal circumstances, since this rarely happens in most professional contexts, this type of repetition should not matter if the participants are not informed of this less-than-natural practice. Instead, from their perspective, the task would simply be yet another one that they were completing as part of their

professional workload. As mentioned, naturalistic tasks can be considered to be somewhere on a continuum between highly controlled and natural conditions, constructed to simulate authentic activities that may or may not involve participants' awareness of exactly what aspects of those activities are being observed and why.

When designing naturalistic studies with naturalistic tasks, serious consideration must be given to the nature of the task and the overarching research question and sub-questions, if the latter have been formulated. A very helpful question for anyone engaging in CTIS research is: Will this study design actually help collect the data needed to answer my research question(s) in a credible and reliable way? As well, if the motivation to participate is not similar to motivators in non-research settings (e.g., course credit for students or remuneration for professionals), then ecological validity and authenticity are potentially compromised, no matter how naturalistic the task might seem. The other criterion for naturalistic tasks to meet the inherent expectations of ecological validity is a setting that makes sense for that task, as discussed in the next section.

In this chapter, we attempt to define the scope of what we understand by naturalistic tasks and explore what they might look like in empirical studies conducted within the CTIS framework. Furthermore, we discuss the settings in which they are or should be used as well as their relevance and the rationale for using naturalistic tasks instead of or in addition to other types of tasks. As with any type of empirical research, the use of naturalistic tasks can raise ethical concerns that could involve much more risk or potential injury to participants than the use of other types of tasks. We close the chapter with a consideration of the emerging challenges in doing research with a commitment to naturalistic tasks and settings in the rapidly evolving field of CTIS.

1. SETTINGS IN NATURALISTIC TASKS

As stressed above, a naturalistic study would take place in the settings where a task is typically performed. Yet some data collection methods might not allow naturalistic observation in authentic settings. In CTIS, these settings might include conference halls, private offices, large company premises, and university classrooms, which may hinder direct observation of fully naturalistic tasks. In contrast, laboratory settings can allow for greater control of stimuli and extraneous variables. Nevertheless, if the setting is very obviously artificial, study participants may not behave in the same way or as well as they would in their natural environment. This change in behavior may be conscious or unconscious and may result from the mere fact that they are in an experimental environment, even if they are performing a familiar and less demanding task. Furthermore, naturalistic settings do not inescapably need to be reproduced in every instance, even if this lack of fully authentic setting is to the detriment of data quality. The extent to which researchers can approach a study environment that is close to an authentic situation will vary depending on the object of the study, the stimuli, and the selected data collection methods.

Much in the same way that the type of task involved in a study falls on a continuum between naturalistic and artificial, so too do the settings in which these tasks are performed. When designing a study, the environment where the task is carried out should be prepared such that

it resembles an authentic setting to the greatest extent possible, provided that the research topic and data collection method allow for this level of control. As Robertson et al. (2018, p.1025) explain:

> […] naturalistic tasks administered in lab-based settings often require imagination and abstract thinking to perform the task in an atypical environment, which limits generalizability to real-world environments […]. To address this limitation, researchers have developed naturalistic tasks that can be performed in realistic environments.

Simulated authentic settings for both rehabilitation and evaluation purposes are used in psychology and neuropsychology (see, e.g., Hudson, 1995 and Robertson et al., 2018), where clinics sometimes build realistic scenarios such as shops, households, offices, and even streets where subjects are asked to perform tasks.

In recent years, immersive technologies have gained ground as a means to simulate a variety of settings for research studies, including virtual reality, augmented reality, and mixed reality. These tools allow the environment to be simulated to varying degrees using technology: virtual reality simulates environments digitally, usually on one (or more) screens; augmented reality, in contrast, aims at making the experience more realistic by adding virtual elements to the real environment; mixed reality, similar to augmented reality, adds virtual elements to the real environment and also makes it possible for users to interact with them (Liberatore & Wagner, 2021). Web-based virtual reality systems do not surround the subject with the virtual reality into which they are introduced, but more recent, head-mounted devices create a more immersive experience for users and can foster a greater sense of engagement in participants (Liberatore & Wagner, 2021). In CTIS, a simulated workplace can vary significantly depending on professional status (e.g., freelancer, in-house, non-professional), activity (e.g., translation, conference or dialogue interpreting, audiovisual translation, respeaking), and context (e.g., large or small companies or events, offices, or institutions).

Translation and interpreting activities always involve a presumed readership or an audience of some type for whom the text or performance is directed. The very presence and awareness of the existence of this audience have an effect on behavior when performing a task. Here, researchers face yet another decision with respect to balancing the advantages and drawbacks of naturalistic settings, deciding between collection of data in the field, the recreation of a workplace in a controlled experimental or laboratory setting, or the use of immersive technologies to augment, mix, or recreate reality in a virtual context.

When collecting data for research on cognitive aspects of interpreting, several factors related to setting merit careful consideration. Many interpreters are likely accustomed to working in different settings with specific equipment or setups. For instance, conference interpreters might work in a booth and use a traditional interpreting console. In an effort to make the task more naturalistic, researchers might ask interpreters to bring their own headset or equipment to emulate a more authentic setting. In a similar vein, community interpreters might work with specific mobile equipment, such that researchers might need to account for familiarity with specific equipment. For sign language interpreting, researchers may need to be aware of their participants' positioning and placement with respect to an audience or camera. In these examples, the research question ought to account for these permutations in order to simulate the nature of the interaction or workplace situation. Other factors must be considered when interpreters work remotely or in virtual environments, including familiarity with specific

interpreting platforms, the physical setting where the task will be performed (e.g., personal office or hub), the physical presence or absence of a boothmate (see Chmiel & Spinolo, 2022), and the supporting technology at hand (e.g., computer-assisted interpreting tools).

A challenging aspect of recreating naturalistic working scenarios for interpreters lies in the interaction with other stakeholders in the communicative event. In some contexts, the participants are the speakers or signers involved in a dialogue or conversation. In other communicative events, there are many more people, including an audience, impromptu speakers, multiple interpreters or colleagues, sound technicians, and other parties such as online attendees. The broad range of potential participants in a communicative event can significantly influence an interpreter's behavior. For example, conference interpreters who monitor audience or speaker reactions may adapt later decisions in their renditions and in their interaction with their boothmates. Likewise, dialogue interpreting involves multiple participants — including the interpreter — co-constructing meaning and the interaction (Wadensjö, 1998). This interactional feature of the task is perhaps the hardest to replicate in a study and is sometimes addressed by resorting to actors or role-players as primary participants that rely on pre-scripted or improvised interaction (see Tiselius & Sneed, 2020) or boothmates (see Chmiel & Spinolo, 2022). When designing a naturalistic study that seeks to replicate these types of settings, researchers need to weigh the benefits and drawbacks of less-controlled naturalistic tasks that occur either in-person or virtually with audiences of various sizes and compositions with the more tightly-controlled experimental setups that provide greater control over variables that might have an effect on interpreters' behavior.

Similarly, translation research must also take into account both the setting and the potential stakeholders. Translators may vary with respect to their typical work settings, which may necessitate a combination of data collection methods, ranging from field observation and remote data collection to laboratory environments. While it may not be possible to recreate the participants' usual work setting in every detail, an attempt can be made to recreate an environment that resembles an actual working space and that can be used for data collection and observation. Workplace studies can include replications of offices with unobtrusive cameras to allow observation, eye tracking systems, or observation rooms with one-way mirrors overlooking the test and focus group rooms (see, e.g., Massey & Ehrensberger-Dow, 2011 and Ehrensberger-Dow & Massey, 2014). However, a translator's working environment comprises not only the physical space in which they work but also, and very importantly, the tools with which they work, ranging from the physical hardware (e.g., keyboards, mouse) to software (e.g., CAT tools, the Internet, digital reference works).

There is thus no one-size-fits-all solution to replicate or simulate the settings in which studies are conducted, but researchers can make conscious decisions in the study design and data collection method(s) that are driven by the specific research question(s) to allow varying degrees of closeness and resemblance to authentic settings. These decisions will also need to account for the influence that the setting and situation can have on translator and interpreter behavior, in line with Muñoz's (2021, p.208) view of cognition being situated and "emerging in situ from the interaction with the body and the environment."

2. PARTICIPANTS IN NATURALISTIC TASKS

A primary objective in designing naturalistic tasks is, presumably, to generate research findings that have relevance and applicability to the translation and interpreting professions or to the training of graduates who will go on to work in those or in related professions. Researchers need to decide the type(s) of participant they will recruit to perform the research tasks. For CTIS research it is not uncommon to recruit translation or interpreting students as convenient proxies for 'professional translators.' However, even the definition of a 'professional translator' is not straightforward and so the first question that arises here is for whom is the student acting as a proxy? A professional translator or interpreter is generally assumed to mean somebody who earns money by translating or interpreting content in business or institutional settings (e.g., governmental or in not-for-profit organizations). Yet a professional translator might also have other jobs (e.g., teaching), which sometimes raises questions about what is required to qualify as a 'professional' in a research study. Factors that have been used to profile informants include years of experience (and percent of full-time equivalent) spent translating or interpreting, the number of words translated per annum, and the percentage of one's income generated through translation activity.

Positioning students as proxies for whatever definition of 'professional' is adopted assumes that those students have similar or equal levels of translation proficiency to that of a professional translator. A tendency is to recruit final year undergraduate or postgraduate students as professional proxies. While some students' language, translation, and tool skills may be excellent, it is probably fair to say that the majority would lack the business acumen and expert knowledge gained by professional translators over time and so cannot be said to be representative of professional practice. Notwithstanding that some students might have some or even considerable professional experience, students, regardless of their stage of study, should be clearly labeled as students and ideally should not be used as proxies who represent professional practice for a naturalistic task setting as this will reduce the relevance of the findings for the profession.

Professionals might be invited to participate in experimental naturalistic tasks, but professionals vary in terms of their language, translation, and subjectmatter competence, not to mention their technical competence. Presenting participants as a 'group of professional translators' is probably too vague. Researchers should endeavor to include, for example, commentary and measures on:

- levels of language skills;
- levels of translation skills;
- (professional) experience, especially in relation to the genre of text and the domain involved in the study; and,
- levels of skill in the tools that are being used.

Regarding language proficiency, the methods for testing this will vary along a spectrum from 'quite loose' (assumption about levels based on profile), through 'loose' (self-reporting on some recognized language competence scale, such as the CEFR), to 'rigorous' (e.g., tried and tested language assessment instruments such as IELTS or TOEFL, to name just two for English). Researchers need to consider how relevant the different skills are (listening, speaking, reading, writing) to the research question(s) and the naturalistic task being set.

Regarding translation skills, as with language, assumptions can be made or proficiency can be tested before including participants in a research study. To measure years of relevant experience, ideally the researcher would consider aggregate experience to account for relevant volunteer activities, part-time professional experience, career breaks etc. Experience with a particular domain and genre should also be considered as should the relevance that has for the naturalistic task being set. Participants for a study involving legal translation, for example, should have extensive experience in that domain – unless, of course, the research question focuses on how inexperienced translators approach a specific legal translation task. A second example is research conducted to compare various aspects of the 'translation' process to 'post-editing' of machine translation. A pitfall here is that some studies do not reflect on the fact that their participants may have years of training and professional experience in 'translation' and little to none in post-editing (or whatever the comparative task is). Thirty minutes of training is not adequate, unless the research question focuses on "what can thirty minutes of training bring to participants who are new to task x?" Researchers should reflect on whether their task allocation and comparison are realistic and fair. Ideally, equal experience and/ or skills in both tasks would be used as selection criteria unless, as stated, the research question focuses on the lack of comparative experience.

Finally, as briefly mentioned, abilities with tools should be factored in where relevant to the research design and task being set. For example, the researcher might think about what impact there may be on the product and process if a participant were tasked with using an unfamiliar tool for the first time. Alternatively, researchers might be interested in the 'learnability' of a specific tool. If this is the research question, obviously it is entirely appropriate, if not even essential, to recruit participants with limited or no experience with a particular tool. If a participant has experience in one CAT tool, what impact might there be if it is assumed that use of a 'similar,' but unfamiliar, tool is required? The researcher might wish to use some of the excellent research-based tools that have been developed over time, such as keystroke loggers and eye trackers. These tools allow valuable data to be collected in a controlled way that is otherwise not possible to collect or observe. Yet using these research-oriented tools may threaten ecological validity if the primary interest lies in natural professional translation processes. If ecological validity is an important requirement, it follows that the task set should involve the tools normally used in professional settings.

If we consider the aim of naturalistic tasks in research to more closely represent tasks that groups of people perform outside the academic environment, then we can also assume that generalizability from that research might be desirable. Generalizability is a huge challenge in research that involves human subjectivity along with varying levels of skill, experience etc., as discussed above. When it comes to research on the cognitive dimensions of translation and interpreting, we know that there is no such thing as 'one correct translation' or one standard process; translators will produce many different versions of a translation, many of which would be accurate and acceptable, using different techniques. Therefore, making textual comparisons and trying to generalize from this is challenging.

A further challenge to generalizability for studies involving naturalistic tasks is the sample size. This is due to a number of factors, including:

- many sample sizes in CTIS tend to be in the tens or twenties, rather than in the hundreds or thousands for various reasons, such as a lack of funding in the field to recruit larger numbers of participants;
- CTIS researchers do not typically work in teams (as engineers do, for example); this limits our ability to scale up analyses that might assist with generalizability; and
- difficulties recruiting a sufficient number of participants with similar backgrounds, experience, and languages.

In addition to the points listed above, when we factor in different language pairs, translation directions, and text type or genre conventions across languages, we are faced with significant challenges for generalizability. We can instead focus on transferability of research findings such that they can potentially inform other contexts or studies. However, to achieve even this for studies with naturalistic tasks, we need to be rigorous in our participant selection and description as discussed earlier.

When considering naturalistic studies, not only do we need to think about participant profiling very carefully, but we also need to consider that what is natural for a student would probably not be natural for a professional. For instance, a student would most likely require and be given more time to produce a translation than a professional. A student may not have built up their resources (corpora, TMs, glossaries) yet, whereas a professional will already have these resources at their disposal. The student's motivation (ideally learning, but also fulfilling course requirements to gain credits) is also different from that of the professional (earning). On the flip side, if we bring a professional into a lab, even if we try to reproduce a 'natural environment,' it will most likely not resemble how they work. Perhaps they work on a laptop, not a desktop, on a Mac, not a PC, with two monitors, not one, with a different type of tool or a different version of it? Maybe their cat is normally curled up in their laps when they work, or they play music in the background? Even their desk and chair will differ from what is available in the lab. Some of these differences are minor and perhaps will not have a big impact on the research task, others are much more likely to do so. We therefore need to understand what we mean by naturalistic, find out what is natural for an individual and, if the research question requires these conditions to be replicated, then we can try to recreate them to the extent possible.

3. ETHICS IN NATURALISTIC TASKS

When designing naturalistic tasks, scholars must be conscious of the research enterprise within which CTIS research is conducted. In many instances, researchers are required to seek approval from various stakeholders prior to conducting studies, including institutional entities that provide resources to do research as well as study participants or respondents. When these studies involve human participants or subjects, one of the most common forms of this type of approval is an ethics review. Institutional review boards or ethics committees typically oversee research in order to protect participants from undue or unwarranted harm while also seeking to mitigate potential risk or harm to the institution (Israel, 2015). The auspices under which research may be conducted and the approval process vary across institutions and countries; however, the overarching tenets of risk mitigation remain central to their work. Researchers should investigate local requirements to ensure that appropriate authorization and protocols

are followed prior to conducting research (Saldanha & O'Brien, 2013). Depending on the perceived or actual risk involved in the study, the approval process can range from a relatively expedited process to a full committee review involving multiple stakeholders. Since the primary focus of this chapter is the design of naturalistic tasks, we will forego a discussion of all the intricacies of ethics review in CTIS research and instead address the ethical considerations related to task specific questions at three interrelated levels: the design and implementation of tasks, the interpretation of task-derived data, and the contexts or settings within which tasks are conducted[1].

A primary concern is the design and implementation of a specific task to collect data in the service of the overarching research question(s). Ethics review committees often evaluate the proposed tasks or data collection methods with an eye toward the level of risk incurred by participants, institutions, and stakeholders in the research process. Determinations of risk do not preclude researchers from conducting research; rather, they require that researchers disclose potential risks and benefits to study participants prior to their involvement in the study. This process, commonly referred to as 'informed consent,' is integral to studies involving human subjects, and is built on the principle of participant autonomy. Through the act and process of informed consent, researchers describe the study in writing, including potential risks and benefits along with an overview of the study procedures, to allow participants to independently determine whether they wish to be involved. In the case of CTIS, this consent typically involves agreeing to complete a series of tasks. In addition, participants are typically given the opportunity to ask additional questions about the study to clarify any elements before voluntarily electing to enroll or to decline.

The consent process is often seen as a hallmark of ethical research, yet the specifics of the process have been called into question when conducting research in the humanities and social sciences (Israel, 2015). More specifically, the legal and regulatory nature of informed consent derives in large part from medical research and as a reaction to historical improprieties involving human participants that forcefully imposed medical interventions on unwilling participants. Naturalistic tasks and social science data collection methods have been argued as being sufficiently distinct from this type of research and of potentially lower risk, thereby challenging the applicability of these regulatory hurdles in all research contexts (Schrag, 2019). In the United States, recent revisions to the Federal Policy for the Protection of Human Subjects (Common Rule) seek to improve the informed consent process, yet questions remain regarding the appropriateness of a single regulatory guideline governing research outside of the biomedical domain (Schrag, 2019; Bell & Wynn, 2020). In CTIS, informed consent has not received much attention in the research literature, although some recent contributions highlight the importance of participants having control over data generated while completing study-specific tasks, including the ability to revoke consent at any point in the process (for instance, see Tiselius, 2021).

An increasingly prominent question in CTIS is whether providing a complete description of the study task as part of the informed consent process could undermine the researcher's ability to observe the phenomena of interest. While observer effects are ostensibly unavoidable in most

[1] There are several overviews of broader questions of research ethics in translation and interpreting studies as well as specific guidance to conduct ethical research in the field, such as Mellinger & Baer (2021) and Saldanha & O'Brien (2013).

research involving participants, researchers ought to acknowledge and take into account these epistemological challenges when analyzing and interpreting data. In this respect, Luhmann's (1990) distinction between first order and second order observations is a useful reminder of the types of observations that can be made. As Baraldi and Mellinger (2016) discuss in a review of Luhmann's work, "first order observations, which posit activities as facts" and "second order observations, which acknowledge the researcher's influence on the observation process," (p.257) and between researchers as observers and researchers as participants. In the first case, "This type of observation situates the researcher as a neutral outside party to a given activity or interaction, though achieving absolute neutrality can be difficult," while in the latter case "the researcher operates from within the social interaction or context" (Baraldi & Mellinger, 2016, p.257).

Observer effects are of particular concern when observing naturalistic behavior during the completion of a task (see, for instance, Jorgensen, 2020; Mellinger, 2020a). One way to overcome this challenge is through the use of partial disclosure or deception during the initial consent process, wherein participants are not told some details of the tasks that they will perform to avoid changing their behavior. Ethics review committees have provisions to allow informed consent to be waived under specific circumstances as well as only partial disclosure of the aims of the study during the informed consent process, often with the proviso that the study's true purpose be revealed upon its completion (Wilson, 2015). However, the utility and appropriateness are often weighed against the ability of participants to make informed decisions about their participation and completion of the study's tasks. Examples exist of CTIS research employing what are known as deception studies (for an overview, see Eyde, 2000 or Hertwig & Ortmann, 2008), particularly involving the role of emotion and affect on tasks. For instance, Rojo and Ramos (2016) provided positive and negative feedback during a translation task to elicit specific emotional responses, which would have otherwise been impossible to ascertain if the entire study design were to be revealed to participants ahead of time. In a similar vein, physiological studies may also require the use of (partial) deception to avoid influencing autonomic responses of participants involved in research on stress, emotion, and ergonomic demands (for an overview of these types of studies, see Gieshoff et al., 2021).

Related task-level ethical considerations are confidentiality and anonymity. In many research designs and task configurations, researchers provide a certain level of reassurance to participants that their identities will not be revealed or discoverable as part of their participation in a study. In the case of confidentiality, their identity is known to the researchers but not revealed, while anonymous participation allows participants to complete tasks without revealing their identity to any of the stakeholders, including the researchers. Naturalistic tasks, particularly those involved in workplace studies, are often difficult to make truly anonymous, insofar as these studies involve direct observation of participants throughout the completion of specific tasks or work. Nevertheless, researchers can protect participants' identity in these settings by assigning participant codes or pseudonyms to collected data and decoupling these identifiers from the actual participant identities (Ehrensberger-Dow, 2014; Ehrensberger-Dow & Hunziker Heeb, 2016).

The same question of confidentiality and anonymity arises when enrolling students as proxies for pseudo-professionals, insofar as researchers will know the identities of these participants. This situation is further complicated by the potentially coercive nature of the recruitment

process, in which researchers may hold additional influence to ask their own students to enroll in the study (Mellinger & Hanson, 2017). A variety of means may potentially mitigate this situation; in addition to guaranteeing the anonymity of student data, researchers may also request that an assistant or colleague approach students as potential participants to lessen the power differential between researcher and participant. The same holds true for the actual data collection; behavior exhibited during tasks may differ if the student–teacher power differential remains during task performance given the potentially vulnerable position of students (Mellinger & Baer, 2021). Another possibility is to use student assignments as a data source after a course has concluded, with the provisos that students have granted permission for their work to be used in this manner and informed consent has been obtained. Researchers may also need to comply with additional regulations and ethics committee guidelines on student participation in research studies, particularly in relation to incentives and course credit.

At the level of data interpretation, additional ethical considerations must be taken into account, particularly when designing naturalistic tasks to elicit data on translation and interpreting. One such ethical tenet is explicit recognition of the relationship and dialectic between participants and the researchers. Researchers are embedded within the observed event, either by means of being present during the task or observed event (Mellinger, 2020a) or as coming from the community being investigated (Cokely, 2005), thereby epistemologically enmeshing the researcher in the interpretation of data. As a result, researchers designing naturalistic tasks must lay bare the links between their own positionality and the object of analysis to recognize the situated nature of this type of research. Exercising reflective practice when analyzing the results of naturalistic tasks allows researchers to mitigate the possibility of their own or their readers' misrepresentation, misinterpretation, or overgeneralization of observed behaviors.

An added dimension of designing naturalistic tasks for CTIS with respect to the relative position of various stakeholders involves the increasingly common practice of involving industry or institutional stakeholders in research projects. These collaborations can be quite fruitful, insofar as they can provide access to hidden or inaccessible data while providing insights on professional translation and interpreting behavior (e.g., Lomeña, 2020; Mellinger, 2020b). This type of research is often referred to as 'engaged research' or 'community-based participatory research' and moves toward the naturalistic end of the continuum of research tasks. With multiple stakeholders or partners involved, important factors arise with respect to the agreements that facilitate this type of research, including which entities have the final say in whether or how specific findings are disseminated as well as the power dynamics between informants and researchers (see Turner & Harrington, 2000). Nevertheless, researchers in partnership with these institutional entities will need to divulge these relationships to limit the potential for conflicts of interest, particularly those that may arise when facing unexpected or less than favorable results. Here, the question of complete reporting comes into play; researchers working with naturalistic tasks must negotiate a matrix of ethical obligations to accurately report the study design and findings while still protecting the identity of participants. Full disclosure of task design and data collection provides sufficient information to understand the project and potentially replicate the study in new environments, while still recognizing the embedded nature of these tasks.

As noted above, naturalistic studies are designed in an effort to understand translation and interpreting situated within authentic settings. Broader ethical questions related to the setting

in which tasks are performed are linked to questions of task design and the interpretation of data, and these considerations are likely to be mutually influential. For instance, the issues of anonymity and confidentiality are linked to various aspects of the study design simultaneously. While participant anonymity may help to mitigate observer effects and potential harm, this practice also helps to reduce risk that may be incurred in their workplace professionally or reputationally. Similarly, institutions involved in or sponsoring research may have a vested interest in maintaining a degree of anonymity or confidentiality to limit risk. However, there is an inherent tension between confidentiality, full disclosure, and reputational risk. Though non-disclosure agreements may help to protect trade secrets or limit the exposure of confidential information, these restrictions may impinge on researchers' abilities to fully describe observed task behavior or study results. Consequently, researchers must be mindful to report potential or perceived conflicts of interest to allow the research community to evaluate the claims based on their merits and the relationships between the various stakeholders.

4. CLOSING REMARKS

In this chapter we have grappled with how to define a naturalistic task, although we would reiterate that the important point is whether a task can actually allow us to collect the data we need to answer our research question(s) in a credible and reliable way. Another challenge is the relevance of the research to the broader context in which translation and interpreting happens. This is especially critical in the fast-moving reality of technologized, multilingual societies that require T&I services; tasks that might be considered naturalistic today may very well be outdated in the near future.

Equally important considerations in study design are data archiving and reuse. Collecting naturalistic task data is laborious, time-intensive, and costly. Researchers may be interested in re-using or re-analyzing data in subsequent studies to maximize limited resources to conduct research. Recent calls in the field for open access data and data repositories tout the benefits of this type of research, yet ethical questions remain. For instance, researchers must be mindful of broad legal requirements, such as the General Data Protection Regulation (GDPR, 2016) in the European Union, which provides individuals with greater control over their personal data that has been collected and how the data are used. In a similar vein, informed consent protocols may have granted researchers with permission to record or access data for a specific purpose, and future use or analysis by the same or different researchers may contravene the granted permissions (see also Mellinger, 2020b). CTIS researchers must weigh setting-specific questions when designing naturalistic tasks, particularly since the decisions made in the initial stages of a research project are likely to have lasting repercussions in the analysis and reporting stages.

To conclude, not every study needs to approach the most 'naturalistic' set up. Everything depends on the research question(s) and the scope of the research project. A researcher is not necessarily obliged to serve the translation and interpreting profession in their research, but the research will ideally have some relevance and applicability to society as a whole. If the work involves a sole researcher with a few months to conduct a study, that person will need to consider what is essential and what might be sacrificed. Research is often about weighing pros

and cons regarding approaches and making decisions based on the research goals and the resources available.

FURTHER READINGS ON NATURALISTIC TASKS

Matusz, P.J., Dikker, S., Huth, A.G., & Perrodin, C. (2019). Are we ready for real-world neuroscience? Journal of Cognitive Neuroscience, 31(3), 327–338.

Mellinger, C.D., & Hanson, T.A. (2022). Considerations of ecological validity in cognitive translation and interpreting studies. Translation, Cognition & Behavior, 5(1), 1–26.

Risku, H., & Windhager, F. (2013). Extended translation: A sociocognitive research agenda. Target. International Journal of Translation Studies, 25(1), 33–45.

REFERENCES

Albl-Mikasa, M., Ehrensberger-Dow, M., Hunziker Heeb, M., Lehr, C., Boos, M., Kobi, M., Jäncke L., & Elmer, S. (2020). Cognitive load in relation to non-standard language input: Insights from translation, interpreting and neuropsychology. *Translation, Cognition & Behavior*, 3(2), 263–286.

Baraldi, C., & Mellinger, C.D. (2016). Observations. In C.V. Angelelli & B.J. Baer (Eds.), *Researching translation and interpreting* (pp. 257–268). Routledge.

Bell, K., & Wynn, L.L. (2020). Research ethics committees, ethnographers and imaginations of risk. *Ethnography*, 24(4) 537–558.

Chmiel, A., & Spinolo, N. (2022). Testing the impact of remote interpreting settings on interpreter experience and performance — Methodological challenges inside the virtual booth. *Translation, Cognition & Behavior*, 5(2), 250–274.

Chou, I., Hu, J., Muñoz, E., & García, A.M. (2021). Discourse-level information recall in early and late bilinguals: Evidence from single-language and cross-linguistic tasks. *Frontiers in Psychology*, 12(757351).

Cokely, D. (2005). Shifting positionality: A critical examination of the turning point in the relationship of interpreters and the Deaf community. In M. Marschark, R. Peterson, & E.A. Winston (Eds.), *Sign language interpreting and interpreter education: Directions for research and practice* (pp. 3–28). Oxford University Press.

Crum, J. (2021). Understanding mental health and cognitive restructuring with ecological neuroscience. *Frontiers in Psychiatry*, 12, 697095.

De Beaugrande, R.A., & Dressler, W.U. (1981). *Introduction to text linguistics*. Longman.

Ehrensberger-Dow, M. (2014). Challenges of translation process research at the workplace. *MonTI*, Special Issue 1, 355–383.

Ehrensberger-Dow, M., & Hunziker Heeb, A. (2016). Investigating the ergonomics of a technologized translation workplace. In R.Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 69–88). John Benjamins.

Ehrensberger-Dow, M., & Massey, G. (2014). Cognitive ergonomic issues in professional translation. In J.W. Schwieter & A. Ferreira (Eds.), *The development of translation competence: Theories and methodologies from psycholinguistics and cognitive science* (pp. 58–86). Cambridge Scholars.

Ehrensberger-Dow, M., & Massey, M. (2019). Translation workplace-based research. In M. O'Hagan (Ed.), *Routledge handbook of translation and technology* (pp. 354–369). Routledge.

Eyde, L.D. (2000). Other responsibilities to participants. In B.D. Sales & S. Folkman (Eds.), *Ethics in research with human participants* (pp. 61–74). APA.

GDPR. (2016). General Data Protection Regulation. Retrieved on 4 November 2024 from https://gdpr-info.eu/

Gieshoff, A.C., Lehr, C., & Hunziker Heeb, A. (2021). Stress, cognitive, emotional, and ergonomic demands in interpreting and translation: A review of physiological studies. *Cognitive Linguistic Studies*, 8(2), 404–439.

Hertwig, R., & Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, 18(1), 59–82.

Hudson, T. (1995). Learning on Easy Street. *Hospitals & Health Networks*, 69(16), 41.

Israel, M. (2015). *Research ethics and integrity for social scientists: Beyond regulatory compliance* (2nd ed.). Sage.

Jorgensen, D.L. (2020). *Principles, approaches and issues in participant observa*tion. Routledge.

Kuznik, A., & Verd, J.M. (2010). Investigating real work situations in translation agencies. Work content and its components. *Hermes*, 44, 25–43.

Liberatore, M.J., & Wagner, W.P. (2021). Virtual, mixed, and augmented reality: A systematic review for immersive systems research. *Virtual Reality*, 25(3), 773–799.

Liu, M., & Chiu, Y.C. (2009). Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgement. *Interpreting*, 11(2), 244–266.

Lomeña Galiano, M. (2020). Finding hidden populations in the field of translating and interpreting: A methodological model for improving access to non-professional translators and interpreters working in public service settings. *FITISPos International Journal*, 7(1), 72–91.

Luhmann, N. (1990). *Die Wissenschaft der Gesellschaft*. Suhrkamp.

Massey, G., & Ehrensberger-Dow, M. (2011). Commenting on translation: Implications for translator training. *Journal of Specialised Translation*, 16, 26–41.

Mellinger, C.D. (2020a). Positionality in public service interpreting research. *FITISPos International Journal*, 7(1), 92–109.

Mellinger, C.D. (2020b). Core research questions and methods. In E. Angelone,

M. Ehrensberger-Dow, & G. Massey (Eds.), *Bloomsbury companion to language industry studies* (pp. 15–35). Bloomsbury.

Mellinger, C.D., & Baer, B.J. (2021). Research ethics in translation and interpreting studies. In K. Koskinen & N.K. Pokorn (Eds.), *Routledge handbook of translation and ethics* (pp. 365–380). Routledge.

Mellinger, C.D., & Hanson. T.A. (2017). *Quantitative research methods in translation and interpreting studies*. Routledge.

Muñoz Martín, R. (2021). Situated cognition. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies* (pp. 207–212). John Benjamins.

Risku, H., Rogl, R., & Milosevic, J. (2020). Researching workplaces. In E. Angelone, M. Ehrensberger-Dow, & G. Massey (Eds.). *Bloomsbury companion to language industry studies* (pp. 37–62). Bloomsbury.

Robertson, K., & Schmitter-Edgecombe, M. (2017). Naturalistic tasks performed in realistic environments: A review with implications for neuropsychological assessment. *The Clinical Neuropsychologist*, 31(1), 16–42.

Robertson, K., Schmitter-Edgecombe, M., Weeks, D., & Pimentel, J. (2018). Naturalistic assessment using a simulated environment: Cognitive correlates and relationship to functional status in individuals with neurologic conditions. *Archives of Clinical Neuropsychology*, 33(8), 1024–1039.

Rojo López, A.M., & Ramos Caro, M. (2016). Can emotion stir translation skill? Defining the impact of positive and negative emotions on translation performance. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 107–130). John Benjamins.

Saldanha, G., & O'Brien, S. (2013). *Research methodologies in translation studies*. Routledge.

Salkind, N.J. (2010). *Sage encyclopaedia of research design*. Sage.

Schrag, Z.M. (2019). Vexed again: Social scientists and the revision of the Common Rule, 2011–2018. *The Journal of Law, Medicine & Ethics*, 47(2), 254–263.

Shreve, G.M. (2006). The deliberate practice: Translation and expertise. *Journal of Translation Studies*, 9(1), 27–42.

Shreve, G.M. (2018). Text linguistics, translating, and interpreting. In K. Malmkjaer (Ed.), *Routledge handbook of translation and linguistics* (pp. 165–178). Routledge.

Teixeira, C.S.C. (2014). *The impact of metadata on translator performance: How translators work with translation memories and machine translation* (Unpublished doctoral dissertation). Universitat Rovira I Virgili.

Teixeira, C.S.C., & O'Brien, S. (2017). Investigating the cognitive ergonomic aspects of translation tools in a workplace setting. *Translation Spaces*, 6(1), 79–103.

Tiselius, E. (2021). Informed consent — An overlooked part of ethical research in interpreting studies. *INContext: Studies in Translation and Interculturalism*, 1(1), 83–100.

Tiselius, E., & Sneed, K. (2020). Gaze and eye movement in dialogue interpreting: An eyetracking study. *Bilingualism: Language and Cognition*, 23(4), 1–8.

Turner, G.H., & Harrington, F. (2000). Issues of power and method in interpreting research. In M. Olohan (Ed.), *Intercultural faultlines: Research models in translation studies I: Textual and cognitive aspects* (pp. 253–265). St. Jerome.

Wadensjö, C. (1998). *Interpreting as interaction*. Routledge.

Wilson, A.T. (2015). Counterfactual consent and the use of deception in research. *Bioethics*, 29(7), 470–477.