

## **Conducting experimental research in audiovisual translation (AVT): A position paper**

**Pilar Orero, Universitat Autònoma, Barcelona; Stephen Doherty, University of New South Wales, Sydney; Jan-Louis Kruger, Macquarie University, Sydney; Anna Matamala, Universitat Autònoma, Barcelona; Jan Pedersen, Stockholm University; Elisa Perego, University of Trieste; Pablo Romero-Fresco, University of Vigo; Sara Rovira-Esteva, Universitat Autònoma, Barcelona; Olga Soler-Vilageliu, Universitat Autònoma, Barcelona and Agnieszka Szarkowska, University of Warsaw**

### **ABSTRACT**

Experimental studies on AVT have grown incrementally over the past decade. This growing body of research has explored several aspects of AVT reception and production using behavioural measures such as eye tracking, as well as venturing into physiological measures such as electroencephalography (EEG), galvanic skin response, and heart rate. As a novel approach to the field of AVT, the experimental approach has borrowed heavily from other fields with established experimental traditions, such as psycholinguistics, psychology, and cognitive science. However, these methodologies are often not implemented with the same rigour as in the disciplines from which they were taken, making for highly eclectic and, at times, inconsistent practices. The absence of a common framework and best practice for experimental research in AVT poses significant risk in addition to the potential reputational damage. Some of the most important risks are: the duplication of efforts, studies that cannot be replicated due to a lack of methodological standardisation and rigour, and findings that are, at best, impossible to generalise from and, at worst, invalid. Given the growing body of work in AVT taking a quasi-experimental approach, it is time to consolidate our position and establish a common framework in order to ensure the integrity of our endeavours.

This chapter analyses problems and discusses solutions specifically related to the multidisciplinary nature of experimental AVT research. In so doing, it aims to set the course for future experimental research in AVT, in order to gain credibility in the wider scientific community and contributes new insights to the fields from which AVT has been borrowing. Its conclusion lays out the foundation for a common core of measures and norms to regulate research in the growing field of AVT.

### **KEYWORDS**

Experimental research, Audiovisual Translation, methodology, eye-tracking, subtitling.

## **1. Introduction**

Audiovisual Translation (AVT) as a field of research is growing exponentially, now also encompassing other fields including Media Accessibility. As with any growth in academic disciplines, levels of complexity are compounded, along with the need to adjust the identity of the field to reflect the state of the discipline overall. Researching AVT, and in particular focusing on its reception, necessitates turning to disciplines outside of traditional translation studies (TS) to establish new interdisciplinary connections and multidisciplinary approaches while becoming transdisciplinary in nature. According to Choi and Pak (2006: 351), "multidisciplinarity draws on knowledge from different disciplines, but

stays within their boundaries.” Interdisciplinarity analyses, synthesises and harmonises links between disciplines into a coordinated and coherent whole. Transdisciplinarity integrates the natural, social and health sciences in a humanities context, and transcends their traditional boundaries.” In this sense, experimental research on AVT has to become a transdisciplinary endeavour; working beyond the boundaries of and integrating approaches from film studies, literary studies, psycholinguistics, cognitive science, and TS. Choi and Pak (*Ibid.*) emphasise that “[t]he objectives of multiple disciplinary approaches are to resolve real world or complex problems, to provide different perspectives on problems, to create comprehensive research questions...”. This emphasis on the need to develop consensus on definitions and guidelines when multiple disciplinary approaches are chosen is an important consideration for all research drawing upon multiple disciplines.

The position articulated in this chapter on experimental AVT research addresses various components of and requirements for conducting experimental research in transdisciplinary teams to tackle complex research questions and to produce empirically-grounded results. This transdisciplinary teamwork has many benefits but also a number of potential pitfalls. On the positive side, it allows to transcend the traditional boundaries of AVT research while drawing on interdisciplinary knowledge and using multidisciplinary methodologies. This allows us to arrive at research questions that are multifaceted and nuanced. On the negative side, these approaches are often met with suspicion by traditional publication outlets in TS, as well as in the journals of other disciplines. This is exacerbated by the fact that technical, authorship, and statistical conventions vary across disciplines, as do experimental protocols, and a common terminology has to be established.

Multidisciplinary research is often strongly encouraged as a requirement for successful proposals to most funding calls. However, we argue that evaluation panels tend to undermine this by penalising multidisciplinary applications that do not align wholly with a particular paradigm. Disciplinary domains are usually classified in a very old-fashioned and compartmentalised manner, such as Social Sciences and Humanities, Physical Sciences and Engineering, and Life Sciences, etc. The same goes for keywords to define proposals and evaluation panels. So, even when transdisciplinary teams are set up with multidisciplinary research methodologies, it remains difficult to draft projects that would fit the requirements of different funding bodies.

Despite these constraints, there has been a rise over the past decade in experimental research in AVT employing a wide variety of methods and technologies from other fields such as linguistics, psychology, cognitive science, media studies, and computer and/or communication science in an interdisciplinary, multidisciplinary and sometimes transdisciplinary fashion. For instance, AVT scholars now regularly employ conventional

questionnaires alongside psychometric methods such as self-rating scales and physiological instruments like eye-tracking, electroencephalography, electrodermal measures and heart-rate monitors. Many interesting findings have resulted from these studies, but the time has come to introduce order to the discipline by establishing standardised experimental protocols and frameworks that will allow to conduct scientifically sound, ethical and replicable studies that will yield much more robust results that can continuously build on our body of knowledge and avoid the interminable reinvention of the experimental wheel and facilitate publication in outlets that will enhance the impact of our work.

This chapter aims to gather some basic principles for experimental research on AVT. The authors have made an effort to share and build on fundamental research principles to agree on a common framework. The discussion provides a critical discussion of various aspects and some recommendations. Section 2 deals with general aspects to be considered when carrying out AVT experimental research: it first describes what an experimental design requires and then delves into specific issues such as ethics and data protection, sampling, and material selection. Section 3 deals with specific research methods such as eye tracking, electroencephalography, psychometrics and electrodermal activity, and concludes with some general recommendations. Section 4 approaches specifically the issue of research publishing and impact.

## **2. General aspects in AVT experimental research**

This section describes the range of general aspects that need to be considered when planning and carrying out AVT research. It first provides a description of the requirements of an experimental design and then moves to address several specific issues inherent in these designs, namely, ethics, sampling, and material selection.

### **2.1. General principles of experimental design and procedure**

There are different research designs available to AVT research. Among these, the experimental design provides a basic model for comparison and replication. The research question is the core of the design of an experiment, and indeed of any research project. Without a well-defined research question that is operationalised properly, an experiment lacks purpose and cannot lead to valid results. Research questions should be clearly defined and based on previous literature and findings (e.g., academic and industry). It is essential to articulate the research question(s) for an experiment, the hypotheses informed by previous research, and the operationalisation of the research question(s) in terms of independent and dependent variables clearly. It is also important to include the justification of statistical methods during the operationalisation.

Experimental design implies the formulation of a question that leads to a hypothesis informed by previous research findings. Subsequently, an experiment is conducted in order to test the hypothesis and either validate or reject it. The data generated in the experiment then has to be analysed and interpreted in the context of the defined hypothesis. As such, an experiment is considered to be a procedure taken to support, refute, or validate a given hypothesis. An informed hypothesis is explicitly stated prior to experimentation and then returned to after data analysis has been completed.

According to Biddix (2009), research questions should be worth investigating, contribute to knowledge and value to the field, improve educational practice, and improve the human condition. Characteristics of a good research question are that the question is feasible, clear, significant, and ethical. Additionally, a good hypothesis must include three components: the variables, the population, and the relationship between the variables.

A pure experiment requires the random assignment of participants to a control and treatment groups in order to identify the effect(s), if any, of a specified treatment by comparison with an appropriate control. Where there conditions are not possible, e.g., in absence of random assignment or a being unable to identify and control for confounding variables, a quasi-experiment is appropriate (also known as natural experiments or field experiments) and should be explicitly stated in the description of the research design.

Due to a plethora of factors (e.g., participant variables, cognitive, linguistic, and sociocultural factors), pure experiments are often difficult to design in AVT. An experiment allows for random assignment of participants to a control or treatment group. Depending on the research question, control groups may also not be possible. While not as rigorous as experiments, quasi-experiments are often a necessary alternative in AVT that allow researchers to assign participants to groups based on characteristics and the factors mentioned above. Quasi-experiments, however, can lack internal validity and run a higher risk of having a more limited, if any, generalisability and replicability. Similarly, case study designs allow researchers even more freedom in the experimental design, but obviously further jeopardise the validity and generalisability of findings. Such approaches may be useful in pilot testing to inform the design of the main experiment.

Mixed-methods research designs have become commonplace in TS research. Such designs combine and triangulate both quantitative and qualitative research to overcome the limitations of each approach on its own. Creswell (2007) provides one of many accessible descriptions of these approaches to research and experiment design. Such resources should be consulted prior to the design of the experiment.

## **2.2. Ethics and data protection**

Ethics refers to norms for conduct that distinguish between acceptable and unacceptable behavior (Resnik 2015). Basic ethical and legal principles underlie all scholarly research and writing to ensure the accuracy of scientific knowledge, to protect the rights and welfare of research participants and to protect intellectual property rights (APA 2010: 11). Researchers follow principles and updates established by their professional associations. These principles include the design and implementation of research involving experimentation, various aspects of scientific misconduct (such as fraud, fabrication of data and plagiarism), regulation of research, protection of the rights of participants such as anonymity, and the protection of vulnerable populations.

The growing interest in reception studies, the widespread use of both behavioural and physiological measures, the growing interest in media accessibility and the consequent involvement of vulnerable audiences in AVT experimentation (e.g. deaf and hard of hearing, blind and visually impaired, elderly, and children) highlight the need for norms for conduct that can guide researchers in actual research situations. Furthermore, applying for ethical approval is becoming an increasingly common step in the execution of AVT research projects (Pérez-González 2014).

Establishing ethically and legally acceptable methods regulating AVT empirical research is mainly needed in order to protect interviewees/participants' rights (see below), prevent falsification of data and modification of results, assure replicability of experiments and increase the responsibility of researchers.

It is particularly important that human rights are safeguarded. These include, according to the 2010 publication of the National Institute of Justice in the United States (Human subject research n.d.):

- Voluntary, informed consent
- Respect for persons: treated as autonomous agents
- The right to end participation in research at any time
- Right to safeguard integrity
- Benefits should outweigh cost
- Protection from physical, mental and emotional harm
- Access to information regarding research
- Protection of privacy and well-being.

For research funded by the European Commission, there are some guidelines to be followed in "The European Code of Conduct for Research Integrity" where advice can be found for human subject research in the social sciences. This often involves surveys, questionnaires, interviews, and focus groups. These are the tools that are typically used also in AVT

experimental research, even though recently physiological measures have been resorted to (eye tracking, electrodermal activity, EEG, heart rate). Adopting and adapting a set of existing norms could be a first step forward to deal consistently with all those practical issues related to actual research situations.

Before an AVT study begins, the researcher should obtain the necessary approvals from the relevant ethics committee at their institution. This typically presupposes the following aspects: participants should give informed consent; data should be anonymised; data should be stored in a secure place for a set period (typically 5 years); privacy, perceived and real benefits from a study, and other relevant considerations should be considered and reported. Publications should report the status of ethics applications and clearance.

One of the critical aspects mentioned in the previous paragraph is data protection, which should be considered while drafting ethical considerations. Data protection is related to the anonymisation of sensitive personal data to protect people taking tests or answering questionnaires. Data protection is also related to storage of gathered data, so it is protected from being used further to its original intention. And finally, data protection is also an issue when communicating with end users and storing their personal contact, such as address, email, and phone number — i.e., sensitive data. More detailed information is available in the EU Regulation (Regulation (EU) 2016/679) but each country has its own laws, and should be consulted.

### **2.3. Participant sampling**

Selecting the number and profile of participants is a critical step in experimental research. The number of participants depends on the type of study, and study design (Guest, Bunce and Johnson 2006; Malterud, Siersma and Guassora 2016). In order to enable a valid statistical comparison of means, it is essential to consider statistical power and effect size. Although mixed-effect modelling makes it possible to control for individual differences and thereby allow valid results from smaller samples, in general, the desired number of participants required to reach statistical power has to be calculated (see Whitley and Ball, 2002). Sample sizes of lower than 25 per group are unlikely to yield statistical power. This sample size (Snijders 2005) assumes a relatively homogenous group and comparable groups. When a degree of variability is present, the sample size should be adjusted upwards.

Given the fact that loss of data (attrition) is commonplace in experimental studies in the field and typically ranges from 20% to 30% (Hennink, Kaiser and Marconi 2016), it is prudent to plan to capture data from participants until the desired number is reached for the study (e.g., recruiting 30 participants per group with the expectation that around 5 will not result in complete data). Attrition rates can vary significantly depending on the

duration and complexity of the task and also on individual differences, but could occur within single experiments due to fatigue, loss of engagement or motivation, and accounting for the temporal aspect of data (see below). Attrition is a bigger factor in longitudinal and repeated measures designs where it is not always possible to get the same participants for all the measures or the full period (Hedeker, Gibbons and Waternaux 1999). In order to assure comparability, participant profiling is essential, and has to be reported. This could include cognitive, linguistic, and other profiling depending on the task.

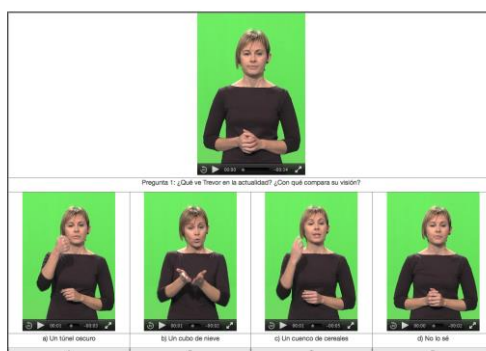
Researchers should make allowances for missing, incomplete or invalid data, resulting from questionnaires where respondents may skip an item or refuse to answer, or calibration issues (as is often the case in eye tracking). The way missing data have been treated needs to be reported as well as how many people were tested, how many data sets were treated as outliers/removed, etc. (cf. APA 2010; McBurney and White 2013).

The participant sample should be identified appropriately and described adequately (APA 2010), including the information on the number of participants, their mean age (and standard deviation), sex, years and type of education, and any relevant details regarding the participant profile, for example hearing (or sight) status, reading proficiency, language skills, and language history. In AVT research, including a section of the questionnaire on TV viewing habits should become a norm. Viewing habits mould the viewers' responses and reactions to any given AVT product (Perego et al. 2016) and should therefore not be missing in AVT-related questionnaires.

Given that participants in many studies on media accessibility are vulnerable users, special care needs to be taken to ensure their fair treatment during the study, and therefore specific ethical procedures should be followed and customised consent forms should be created when conducting research with people with sensory impairments (for a recent example cf. UCLA OHRPP 2016). In the EU public document "How to complete your ethics self-assessment" (EU Guidance 2018: 6), special attention is paid to children and vulnerable participants. Regarding minors, details of the age range are requested, as well as information about the assent procedures and parental consent, about the steps taken to ensure the welfare of minors, and a clear justification of how minors were involved. The form also requests researchers to provide additional details for vulnerable individuals or groups and demonstrate that they have ensured that participants have a fully informed understanding of the implications of participation. These details include the type of vulnerability and details of recruitment, inclusion and exclusion criteria and informed consent procedures.

For instance, Deaf participants who use sign language should be offered sign language interpreting during the study. Also in order to avoid test pollution in the communication with users, from a written to oral language

and interpretation, a sign language questionnaire may be advisable. Figure 1 depicts the SL questionnaire prepared for the HBB4ALL project where sign language users are offered a multiple choice sign language questionnaire.



**Figure 1. Sign language questionnaire developed in EU project HBB4ALL.**

We acknowledge the fact that vulnerable participants are often difficult to recruit, therefore details on the recruitment process should be reported in the paper, also to prevent the frequent reviewer criticism regarding the sample size. However, a critical number of participants with impairments should always be reached in any AVT research experiment to be considered valid and reliable, and therefore publishable.

## 2.4. Materials

The choice of materials will depend on the type of study, its design and research questions, and should also take into consideration copyright issues. Ideally, audiovisual materials will be as authentic as possible, ensuring ecological validity. If fragments of longer stretches of videos are used, care needs to be taken for the clips to be self-contained. If various fragments are compared, they should be similar in terms of complexity, speech rate, genre, etc. so as not to create confounding variables.

Depending on the aim of the test, the length of the clip or clips needed will vary. For instance, for studies testing immersion, it is recommended that complete texts be used where possible, or at least self-contained longer clips. Similarly, benefits of an AVT mode such as subtitling for comprehension, learning or other positive outcomes should be verified through replication as well as longitudinal studies. It should also be kept in mind that a period of acclimatisation may be required in order to measure particular effects, which precludes the use of very short clips. If shorter clips or fragments are used, a large number of these fragments or clips are typically needed for the sake of robustness of results.

When reporting on the experiment, the material should be described in detail, including information on the number and duration of the clips, the original language, the genre, the type of AV translation used in the study (subtitles, dubbing, voice-over, audio description, etc.). In studies on



subtitling, the presentation rate of subtitles should be reported as well as the means by which that was calculated. Furthermore, whenever possible, it is useful to report linguistic data of both the source and the target dialogues (i.e., overall number of words and characters, type/token ratio, mean sentence length, etc.). This will contribute to determine the degree of linguistic complexity of the dialogues (Li 2000; Perego et al. 2016; Szmrecsányi 2004). Due to the multiple channels involved in audiovisual texts, it is important to describe the nature of the information presented visually and auditorily, as well as the density of information (i.e. how much competition for either visual or auditory information a particular film, scene, clip or frame contained). If material was manipulated for experimental purposes, the nature of the manipulation has to be documented.

### **3. Research methods**

To carry out empirical research, various methods and tools can be used. It is essential to select the appropriate tools to answer the specific research question of the study. Broadly speaking, studies that investigate the reception or processing of AVT products can make use of either offline or online measures. Offline measures include self-reported cognitive effort scales, presence or transportation scales, comprehension or retention questionnaires, narrative reports, interviews or focus groups. These measures are commonly post-hoc measures used directly after a participant has been exposed to a text. Online measures allow the researcher to collect data while the participant is processing the text and include eye tracking, EEG, galvanic skin response, and heart rate, among others. These measures will be defined briefly followed by a description of their relative strengths and weaknesses and recommendations for use.

#### **3.1. Eye tracking**

For eye tracking studies, refer to guidebooks such as Holmqvist et al. (2011) or Liversedge, Gilcrest and Everling (2011). For specific applications in AVT research, see Doherty and Kruger (2018), Kruger and Doherty (2016) and Kruger (forthcoming), as well as the previous chapter in this book.

In a paper using eye-tracking methodology, the following data are typically reported: type of eye tracker, sampling frequency, software (Tobii Studio, SMI BeGaze), settings used in the algorithms for event detection (e.g. type of event detection such as dispersion or velocity based, minimum duration for fixation detection is usually around 75 milliseconds (Tobii) and 80 milliseconds (SMI) with a maximum dispersion of around 100 pixels in dispersion-based algorithms) and calibration protocol used. Tracking ratio is important and participants with a tracking ratio of below 90% to 95% should probably be discarded unless otherwise justified (other software uses different terminology, e.g. sample rate in Tobii; some studies use a variety of system-dependent and task-dependent measures to provide a

more robust measurement of eye tracking data quality, e.g., Hvelplund 2011; Doherty 2012).

Typical eye tracking measures used in AVT include mean fixation duration, first fixation duration, number of fixations, dwell time, percentage dwell time, gaze shifts between the subtitle and the image (also referred to as 'deflections'), blink rate and blink count. For AVT specifically, the Reading Index for Dynamic Texts (RIDT) measures degree of processing rather than simply attention to subtitles (Kruger and Steyn 2014). It is "a product of the number of unique fixations per standard word in any given subtitle by each individual viewer and the average forward saccade length of the viewer on this subtitle per length of the standard word in the text as a whole" (Kruger and Steyn 201: 110). A higher RIDT score therefore indicates a higher reading load. Pupil diameter or pupillometry is typically not a useful measure in the context of video due to changes in luminosity, as well as changes in pupil shape as the eye explores various parts of the screen.

Eye tracking is a very useful tool in experimental research in AVT to quantify the attention to and attention distribution between various parts of the screen, as well as to gain an understanding of the nature of the processing. Although heat maps and focus maps give useful qualitative data as well as powerful visualisations of gaze data, they should be used mainly to identify trends that can be investigated quantitatively by looking at fixation data.

### **3.2. Electroencephalography (EEG)**

EEG is a relatively new measurement in the context of translation research and due to the volume and complexity of the data it has to be approached with caution and preferably in collaboration with experts from the field of cognitive science. With the availability of affordable devices such as the Emotiv EPOC+ headset, this type of methodology is becoming more accessible. When reporting EEG data, it is essential to ensure that established protocols are applied for artifact rejection to remove noisy EEG signals and to transform accepted trials. Very little work has been done to date to validate different EEG measures for use in AVT research. It is not recommended to use proprietary software such as that supplied by Emotiv since the manufacturers do not share the algorithms used for data processing, making it impossible to verify the calculations.

Typical measures used in other disciplines such as psychology include alpha and theta power to measure variations in cognitive load, with signals collected in the central, occipital, temporal and parietal regions (see Gerlic and Jausovec 1999; Antonenko et al. 2010; Klimesch et al. 1998; Foxe and Snyder 2011). Beta coherence between prefrontal and posterior regions has also been used as a measure of immersion in the fictional world of film by Kruger, Soto-Sanfiel, Doherty and Ibrahim (2016), based on the work of Reiser Schuller, Weiss, Fink, Rominger, and Papousek (2012) who use state-dependent decreases or increases of EEG coherence between

prefrontal and posterior cortical regions to determine whether these differences indicate a mechanism for modulating the impact of social-emotional information on an individual.

An accessible introduction to event-related potentials can be found in Luck (2014) in addition to a plethora of open-sources resources and toolkits such as EEGLAB (Delorme and Makeig 2004).

### **3.3. Galvanic skin response and heart rate**

Research in AVT and media accessibility using psychophysiological measures is still scarce (Ramos Caro 2015, 2016), but there is some evidence in media research (Ravaja 2004) and also, to a much lesser extent, in interpreting (Kurz 2002).

Emotions can be measured as physiological responses following the activation of the sympathetic nervous system, which alter sweating and heart rate, among other effects. Sweating alters skin conductivity, which can be easily measured using electrodes on hands and fingers. Two of the measures used are electrodermal activity (EDA), also known as galvanic skin response (GSR), and heart rate (Cowley et al. 2016). These measures have been tested in recent studies for their capacity to account for emotional states induced by films (Bos et al. 2013; Brumbaugh et al. 2013; Codisoti, Surcinelli and Baldaro 2008; Fernández et al. 2012). They can be complemented by recordings which capture facial expressions and vocal utterances (O'Hagan 2016), and are used to interpret the subjects' emotional arousal.

### **3.4. Recommendations**

Due to the fact that none of these measures are without limitations, and in order to arrive at robust and replicable results, it is recommended that the data from different measures have to be triangulated. Online measures like eye tracking and EEG can be supplemented by offline measures like post-hoc self-report scales, comprehension and recall tests, or other offline measures like interviews. Such triangulated data also provide a more comprehensive picture of the phenomenon investigated. In view of the multidisciplinary nature of such experiments, it is important to assemble a team of experts from adjacent fields like psychology, cognitive science, cognitive film studies, and educational psychology, as well as members with appropriate knowledge of statistics.

Where possible, it is advisable to use previously validated instruments or questionnaires. Such instruments can be modified if required and justified. Continuous recreation of instruments and questionnaires can lead to an inability to reproduce results and generalise findings as well as a great deal of time and resources spent on creating an instrument or questionnaire that will only be used once. In all cases, the basic psychometric properties of all

instruments, questionnaires and other items of measurement should be reported, including reliability, validity, scales, etc. (refer to American Psychological Association's guide in American Educational Research Association, American Psychological Association, Joint Committee on Standards for Educational and Psychological Testing (U.S.), National Council on Measurement in Education, 2014).

In empirical research, a huge amount of data is often generated. While qualitative data such as heat maps and scanpaths are useful to visualise the results or to trace some initial patterns in the data, researchers in empirical AVT studies need to examine the numerical data using proper quantitative analysis.

Research data obtained in an empirical AVT study should be analysed using appropriate statistical analyses and models. Eye tracking and EEG data, for example, are seldom distributed normally, which means that they cannot be tested using parametric tests like t-tests and ANOVAs unless the data are transformed. When reporting the results, apart from stating how the data were processed (e.g. accounting for normalising, outliers), it is important to report what statistical tests and measures were used, the statistical significance (p value and the significance threshold) as well as effect sizes to demonstrate practical significance. Authors should also state which statistical software was used and which version. Refer to American Psychological Association's guide to reporting statistical findings (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008).

#### **4. Research publishing and impact**

Interesting as the results of an empirical study may be, it is essential for the sake of advancing the discipline to disseminate the information in peer-reviewed publications and also to consider impact. These two aspects—publication in AVT and impact—are discussed in the next paragraphs.

##### **4.1. Publishing research in AVT**

The very nature of AVT makes it a complex field of research that necessitates transdisciplinary, interdisciplinary and multidisciplinary approaches. Although AVT scholars have been aware of this since the origins of AVT research, early approaches to its studies tended to be mono-disciplinary: AVT was mainly studied either from a linguistic or from a translational perspective (think, for example, of the plethora of case studies on the rendering of specific linguistic issues into a target language, such as culturally marked term, forms of address, swear words, etc.). This made it easy to choose the right publication avenue – translation journals.

As mentioned above, the experimental approach to AVT has been learning from and adapting methodologies and technologies from other fields. This

has changed the nature of AVT research and publications, and it is making the selection of the right publication outlet more and more difficult. Authors are often uncertain where to submit their papers and editors often reject papers whose perspective does not fit into their journal's scope. In fact, the borderline nature of some manuscripts that cut across different disciplines makes them unfit for very specialised journals. To date, there is no specific AVT journal, although some translation journals tend to host AVT papers more often than others (e.g. *JosTrans*, *Perspectives*, *Target*, *The Translator*, and *Across Languages and Cultures*). The creation of a specialised indexed journal of AVT would provide a useful avenue for intradisciplinary work. However, in order to advance the discipline and also contribute to other disciplines, it is essential for AVT scholars to pursue other avenues open to transdisciplinary and multidisciplinary approaches. It is only when an AVT study can compete on equal footing with other studies in highly rated cognitive science, computer science, psychology (including educational psychology and media psychology) journals, that the discipline will have matured.

Contributions on AVT have overall grown spectacularly over the years. According to BITRA, until 1980 there were only 78 contributions devoted to AVT (1.3% of the total). The database yields 134 results for the years comprised between 1981 and 1990 (1.8% of the total for that period). The 1991-2000 period represents a turning point in AVT research with 734 outputs (4% of the total). The high productivity (1,789) observed for the first decade of the 21<sup>st</sup> century (2001–2010) seems to be a symptom of the consolidation of this subfield within TS (taking 6.7% of the cake), while the last period available at the time of writing this article (from 2011 to mid-2016) confirms this exponential growing trend, with already 937 contributions (9.8% of the total for that period). As far as methodologies are concerned, 1.9% of all the AVT contributions in the database are somehow linked to the label 'experimental' or 'reception.'

Another aspect worth looking at and especially relevant in terms of citations, impact and research assessment is the degree of collaboration among scholars. Co-authorship in AVT contributions reaches 19.1% on average, which is above the figure in TS (15.8%). Almost 75% of the whole AVT production concentrates in the last 16 years which means that it is a relatively new research area within TS, since only 57% of the whole TS production concentrates in this period. However, this growth has not been followed by the consolidation of its own space within TS in terms of impact and visibility.

#### **4.2. Measuring impact**

Impact is a loaded term with different meanings and interpretations, depending on the context and country in which it is used. According to the Economic and Social Research Council (2016), academic impact is defined as "the demonstrable contribution that excellent social and economic

research makes in shifting understanding and advancing scientific, method, theory and application across and within disciplines.” This is often referred to as ‘contribution to knowledge’ and it may be measured at an individual level and with regard to publications through author-level metrics such as the *h index* or through the impact factors of scientific journals. As mentioned in the previous section, the bias found in the literature on AVT towards book chapters over journal articles works against the visibility of the field in terms of academic impact, given that many of the measures assessing impact through citations do not account for monographs and book chapters.

In this sense, altmetrics would be a good alternative in AVT research assessment to complement traditional, citation-based metrics, since they can include citations on Wikipedia and in public policy documents, discussions on research blogs, mainstream media coverage, bookmarks on reference managers like Mendeley, and mentions on social networks such as Facebook, LinkedIn or Twitter. Altmetrics consist of records of attention, by indicating how many people have been exposed to and engaged with a scholarly output, measures of dissemination and indicators of influence and impact, since some of the data gathered can signal that a given study is changing a field of study or is having a tangible effect upon a given sector of society.

As opposed to academic impact, social impact may be defined as “an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia” (HEFCE 2016). Impact is not regarded here as a contribution to knowledge but rather as the result of non-academic engagement. It is not the process by which research is disseminated either, but its outcome in terms of change, effect or benefit. In countries such as the UK, this notion of impact is now as important as academic impact and has become an inescapable requirement for PhD scholarships, funding and academic promotions. Impact is assessed quantitatively in terms of reach (the size or profile of an audience or an institution benefiting from research, economic indicators, attendance figures, etc.) and significance (personal testimonies, evidence of uptake by external organisations, partnership agreements, inclusion in policy documents, etc.). Although the emphasis is placed on the social outcome, impact must always be underpinned by high-quality research whose findings motivate the effect on non-academic partners.

There is merit in encouraging researchers to leave their ivory tower and engage with society. However, AVT researchers should also be aware of the potential risks involved in working towards a notion of impact that favours short-term findings over long-term results, collaboration with large and powerful partners over small audiences and institutions, and collaborative work with impactful disciplines rather than conceptual work in areas whose non-academic impact may be difficult (or take longer) to obtain. Admittedly, the latter issue does not affect experimental studies on AVT, which are well

positioned with regard to this idea of impact (REF 2014). AVT reception studies, for example, by definition engage with users and, as has been the case with live subtitling (Romero-Fresco 2016) and remote accessibility (Saks and Orero 2015), they often inform national and international policy documents and guidelines. However, AVT researchers are advised to avoid the risks of taking 'shortcuts to impact', such as certain types of commissioned research that may compromise their freedom and independence and where the end result (the social impact) may be seen to shape the research study. It should be possible to keep social impact as a potential and organic outcome of experimental research in AVT while, at the same time, upholding the requirements for scientific rigour described in this chapter.

## **5. Conclusions**

This chapter considered a number of aspects that are fundamental to experimental research in AVT. It is the belief of the authors that experimental research in AVT has the potential to elevate the field into a truly transdisciplinary, interdisciplinary and multidisciplinary endeavour that not only draws on other disciplines, but that strengthens other disciplines and expands our knowledge base in the humanities and the sciences. However, taking this route does require a commitment from researchers in the field to go the distance in terms of scientific rigour, and the application of ethically and legally acceptable research methods. As a first position paper on experimental research in AVT, this chapter therefore lays the foundation for a common core of measures and norms to regulate research in this field and to establish it as an important disciplinary area that will become an important contributing member of the broader scientific community and not a passive user of other disciplines.

This article was written with the overall aim of consolidating the experimental methodology in AVT. The interest is not to restrict research avenues and approaches but the opposite: to encourage creative and original research questions. The article also opens the door to research on hybrid AVT modalities, media formats, and service production and delivery. Mixing subtitles with language technologies for its delivery, applying easy-to-read to any existing service as with easy to read audio description. The manual production, semiautomatic, or automatic production of any AVT modality. Generating the services on an individual or collective way, along quality control, and finally media in its many formats.

This should not be considered an act of breaking ranks with mainstream Translation Studies, but as an evolutionary step towards responsible empirical research that will allow valid, generalisable and replicable conclusions to set the course of future developments in the field. The guidance provided in this chapter is by no means exhaustive, and hopefully it will become redundant in the near future as any true standard should.

## References

- **American Educational Research Association, American Psychological Association, Joint Committee on Standards for Educational and Psychological Testing (U.S.), National Council on Measurement in Education.** (2014). *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association.
- **APA Publications and Communications Board Working Group on Journal Article Reporting Standards** (2008). "Reporting standards for research in psychology: why do we need them? What might they be?" *The American Psychologist* 63(9), 839-851.
- **Antonenko, Pavlo; Paas Fred Roland Grabner and Tamara van Gog** (2010). "Using electroencephalography to measure cognitive load." *Educational Psychology Review*, 22, 425-438.
- **APA (American Psychological Association)** (2010). *Publication Manual of the American Psychological Association*. 6<sup>th</sup> edition. Washington: APA.
- **Biddix, Patrick J.** (2009). "Writing Research Questions." *Research Rundowns* <https://researchrundowns.com/intro/writing-research-questions/> (consulted 02/12/2016).
- **Bos, Marieke; Jentgens, Pia; Beckers, Tom, and Kindt, Merel** (2013). "Psychophysiological response patterns to affective film stimuli." *PloS One*, 8(4), e62661, 1-8.
- **Brumbaugh, Claudia; Kothuri, Ravi; Marci, Carl; Siefert, Caleb and Pfaff, Donald** (2013). "Physiological correlates of the big 5: autonomic responses to video presentations." *Applied Psychophysiology and Biofeedback*, 38(4), 293-301.
- **Choi, Bernard, C. and Pak, A.W.** (2006). "Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness." *Clinical and Investigative Medicine*, 29(6), 351-364.
- **Codispoti, Maurizio ; Surcinelli, Paola and Baldaro, Bruno** (2008). "Watching emotional movies: affective reactions and gender differences." *International Journal of Psychophysiology*, 69(2), 90-95.
- **Cowley, Benjamin et al** (2016). "The psychophysiology primer: a guide to methods and a broad review with a focus on human-computer interaction." *Foundations and Trends in Human-Computer Interaction*, 9, (3-4), 150-307.
- **Creswell, John** (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. 4th edition. London: Sage.
- **Delorme, Aranud and Makeig, Scott** (2004). "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis." *Journal of Neuroscience Methods*, 134(1), 9-21.
- **Doherty, Stephen** (2012). Investigating the effects of controlled language on the reading and comprehension of machine translated texts: A mixed-methods approach (Doctoral dissertation). Dublin City University, Dublin, Ireland.
- **Doherty, Stephen and Kruger, Jan-Louis** (2018). "A systematic review of the eye tracking measures used in empirical research on subtitling and captioning." Dwyer,



Tessa; Perkins, Claire; Redmond, Sean and Sita, Jodi (eds.), *Seeing into Screens: Eye Tracking and the Moving Image*. London: Bloomsbury.

- **ESRC (Economic and Social Research Council)** (2016). "What is impact?" <http://www.esrc.ac.uk/research/impact-toolkit/what-is-impact/> (consulted 04.05.2018).
- **EU Guidance** (2018) 'How to complete your ethics self-assessment.' [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/ethics/h2020\\_hi\\_ethics-self-assess\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf) (consulted 04.05.2018)
- **Fernández, Cristina et al** (2012). "Physiological responses induced by emotion-eliciting films." *Applied Psychophysiology and Biofeedback*, 37(2), 73–79.
- **Foxe, J. John and Snyder, Adam C.** (2011). "The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention." *Frontiers in Psychology*, 2 (154), 1-13.
- **Gerlic, Ivan and Jausovec, Norbert** (2001). "Differences in EEG power and coherence measures related to the type of presentation: text versus multimedia." *Journal of Educational Computing Research*, 25(2), 177-195.
- **Guest, Greg; Bunce, Arwen and Johnson, Laura** (2006). "How many interviews are enough? An experiment with data saturation and variability." *Field Methods*, 18(1), 59-82.
- **HEFCE (Higher Education Funding for England)** (2016). REF Impact. Retrieved <http://www.hefce.ac.uk/rsrch/REFImpact/> (consulted 04.05.2018)
- **Hedeker, Donald; Gibbon, Robert and Waternaux, Christine** (1999). "Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups." *Journal of Educational and Behavioral Statistics*, 24(1), 70-93.
- **Hennink, Monique; Kaiser Bonnie and Marconi, Vincent** (2016). "Code saturation versus meaning saturation: how many interviews are enough?" *Qualitative Health Research* 27(4), 1-18.
- **Holmqvist, Kenneth et al** (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press.
- **Human subject research** (n.d.) Wikipedia.
- [https://en.wikipedia.org/wiki/Human\\_subject\\_research](https://en.wikipedia.org/wiki/Human_subject_research) (consulted 15.05.2016).
- **Hvelplund, Kristian** (2011). *Allocation of Cognitive Resources in Translation: An Eye-tracking and Key-logging study* (Doctoral dissertation). Copenhagen Business School, Copenhagen.
- **Klimesch, Wolfgang; Doppelmayr, Michael; Russegger, Harald; Pachinger, Thomas and Schwaiger, J.** (1998). "Induced alpha band power changes in the human EEG and attention." *Neuroscience Letters*, 244, 73-76.
- **Kruger, Jan-Louis** (forthcoming). "Eye tracking in audiovisual translation research." Luis Pérez-González (ed.). *The Routledge Handbook of Audiovisual Translation Studies*. London: Routledge.
- **Kruger, Jan-Louis and Steyn, Faans** (2014). "Subtitles and eye tracking: reading and performance." *Reading Research Quarterly*, 49(1), 105–120.

- **Kruger, Jan-Louis; Soto-Sanfiel, Maria T.; Doherty, Stepehn and Ibrahim, Ronny.** (2016). "Towards a cognitive audiovisual translatology: subtitles and embodied cognition." Ricardo Muñoz (ed.). *Reembedding Translation Process Research*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 171-193.
- **Kruger, Jan-Louis and Doherty, Stephen** (2016). "Measuring cognitive load in the presence of educational video: towards a multimodal methodology." *Australasian Journal of Educational Technology*, 32(6), 19–31.
- **Kurz, Ingrid** (2002). "Physiological stress responses during media and conference interpreting." Giuliana Garzone and Maurizio Viezzi (eds.), *Interpreting in the 21st Century* 295-202). Amsterdam: Benjamins.
- **Li, Yili** (2000). "Linguistic characteristics of ESL writing in task-based e-mail activities." *System*, 28, 229-245.
- **Liversedge, Simon Gilchrist; Iain and Everling, Stefan** (eds.) (2013). *The Oxford Handbook of Eye Movements*. Oxford: Oxford University Press.
- **Luck, Steven J.** (2014). *An Introduction to the Event-related Potential Technique*. 2nd edition. Cambridge, MA: MIT Press.
- **Malterud, Kirsti; Siersma, Volkert and Guassora, Ann** (2016). "Sample size in qualitative interview studies: guided by information power." *Qualitative Health Research*, 26(13), 26: 1753-1760.
- **O'Hagan, Minako** (2016). "Game localization as emotion engineering: methodological exploration." Minako O'Hagan and Qi Zhang (eds.). *Conflict and Communication: A Changing Asia in a Globalising World*. New York: Nova Publishers, 81-102.
- **Pérez-González, Luis** (2014). *Audiovisual Translation: Theories, Methods and Issues*. London: Routledge.
- **Ravaja, Niklas** (2004). "Contributions of psychophysiology to media research: review and recommendations." *Media Psychology*, 6, 193-235.
- **Ramos Caro, Marina** (2015). "The emotional experience of films. Does audio description make a difference?" *The Translator*, 21(1), 68-94.
- — (2016). *La traducción de los sentidos*. Munich: LINCOM.
- **Reiser, Eva M.; Schulter, Günther; Weiss, Elisabeth M.; Fink, Andreas; Rominger, Christian and Papousek, Ilona** (2012). "Decrease of prefrontal–posterior EEG coherence: loose control during social–emotional stimulation." *Brain and Cognition*, 80, 144–154.
- **REF (Research Excellence Framework)** (2014). "Media for all: Live Subtitling for Deaf and Hard of Hearing People Around the World." <http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=20470> (consulted 15.05.2016).
- **Regulation** (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 <http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679> (consulted 15.05.2016).

- **Resnik, David B.** (2015) "What is Ethics in Research & Why is it Important?" <http://www.niehs.nih.gov/research/resources/bioethics/whatis/> (consulted 04.05.2018)
- **Romero-Fresco, Pablo** (2016). "Accessing communication: The quality of live subtitles in the UK." *Language & Communication*, 49, 56–69.
- **Saks, Andrea and Orero, Pilar** (2015). *FSTP-AM Guidelines for accessible meetings*. Geneva: ITU. [http://www.itu.int/dms\\_pub/itu-t/opb/tut/T-TUT-FSTP-2015-AM-PDF-E.pdf](http://www.itu.int/dms_pub/itu-t/opb/tut/T-TUT-FSTP-2015-AM-PDF-E.pdf) (consulted 04.05.2018).
- **Snijders, Tom** (2005). "Power and sample size in multilevel linear models." Brian S. Everitt, & David C. Howell (eds.), *Encyclopedia of Statistics in Behavioral Science* Vol. 3. Chicester: Wiley, 1570-1573.
- **Szmrecsányi, Benedikt** (2004). "On operationalizing syntactic complexity." Gérard Purnelle; Gérard Fairon and Anne Dister (eds.), *Le Poids des Mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis* (pp. 1032-1038). Louvain-la-Neuve, Belgium: Presses Universitaires de Louvain.
- **The European Code of Conduct for Research Integrity** (2017). [http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics\\_code-of-conduct\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf) (consulted 04.05.2018).
- **UCLA OHRPP (UCLA Office of the Human Research Protection Program)** (2016). "Guidance: Research Involving Visually and/or Hearing Impaired Participants or Participants Who Are Illiterate." [http://ora.research.ucla.edu/OHRPP/Documents/Policy/9/Visually\\_Impaired.pdf](http://ora.research.ucla.edu/OHRPP/Documents/Policy/9/Visually_Impaired.pdf) (consulted 04.05.2018)
- **White, Theresa and McBurney, Donald** (2013). *Research Methods*. 9<sup>th</sup> edition. Belmont: Wadsworth.
- **Whitley, Elise and Ball, Jonathan** (2002). "Statistics review 4: sample size calculations." *Critical Care*, 6(4), 1.

## Biographies

**Pilar Orero**, works at Universitat Autònoma de Barcelona where she leads research projects on media accessibility. She participates in standardisation agencies such as Un ITU and ISO, and the Spanish national agency AENOR.



Email: [pilar.orero@uab.cat](mailto:pilar.orero@uab.cat)

**Stephen Doherty** is Senior Lecturer in the University of New South Wales, Australia, where he directs the Language Processing Lab. His research is based in the interaction between language, cognition, and technology. His current work investigates human and machine language processing, with a focus on psycholinguistics and language technologies.



Email: [s.doherty@unsw.edu.au](mailto:s.doherty@unsw.edu.au)

**Jan-Louis Kruger** is Head of the Department of Linguistics at Macquarie University in Sydney where he also teaches in AVT. His main research interests include studies on the reception and processing of audiovisual translation products including aspects such as cognitive load, comprehension, attention allocation, and psychological immersion.



Email: [janlouis.kruger@mq.edu.au](mailto:janlouis.kruger@mq.edu.au)

**Anna Matamala** is an associate professor at Universitat Autònoma de Barcelona, where she leads TransMedia Catalonia research group. She is involved in media accessibility research projects and standardisation work. Her research interests are media accessibility, audiovisual translation and applied linguistics.



Email: [anna.matamala@uab.cat](mailto:anna.matamala@uab.cat)

**Jan Pedersen** is Associate Professor and Director of the Institute for Interpreting and Translation Studies, and Deputy Head of the Department of Swedish and Multilingualism at Stockholm University, where he researches and teaches audiovisual translation. He has worked as a

subtitler for many years and is the former president of ESIST, and Associate Editor of *Benjamins Translation Library*.



Email: [jan.pedersen@su.se](mailto:jan.pedersen@su.se)

**Elisa Perego** is an associate professor at the University of Trieste. Her research interests and publications lie mainly in the field of audiovisual translation, AVT accessibility and reception, and the use of eye tracking methodology in AVT research. She is currently the coordinator of the European project ADLAB PRO (2016-2019) on audio description.



Email: [eperego@units.it](mailto:eperego@units.it)

**Pablo Romero-Fresco** is a Ramón y Cajal researcher at Universidade de Vigo (Spain) and Honorary Professor of Translation and Filmmaking at the University of Roehampton (UK). He is the author of the books *Subtitling through Speech Recognition: Respeaking* (Routledge) and *Accessible Filmmaking* (Routledge) and leader of the research centre GALMA (Galician Observatory for Media Accessibility), for which he is coordinating the EU-funded projects Media Accessibility Platform and ILSA (Interlingual Live Subtitling for Access).



Email: [promero@uvigo.es](mailto:promero@uvigo.es)

**Sara Rovira-Esteva** has a Ph.D. in Translation Studies. She lectures in Mandarin Chinese and Chinese-Spanish Translation at the Autonomous University of Barcelona (UAB). She is currently Research Coordinator at the UAB Department of Translation and Interpreting and East Asian Studies.

Her research interests include audiovisual translation, Chinese linguistics and bibliometrics.



Email: [sara.rovira@uab.cat](mailto:sara.rovira@uab.cat)

**Olga Soler-Vilageliu** is a tenured Lecturer at the Departament de Psicologia Bàsica, Evolutiva i de l'Educació at Universitat Autònoma de Barcelona, where she teaches Psychology of Language to undergraduate students of Speech Therapy and Psychology. Her main interest in research is language processing, and she is currently involved in projects on literacy learning and media accessibility.



Email: [olga.soler@uab.cat](mailto:olga.soler@uab.cat)

### **Agnieszka Szarkowska**

Agnieszka Szarkowska is Research Fellow at the Centre for Translation Studies, University College London, and Assistant Professor in the Institute of Applied Linguistics, University of Warsaw. She researches subtitling and audio description.



Email: [a.szarkowska@uw.edu.pl](mailto:a.szarkowska@uw.edu.pl)