

CHAPTER VIII

Methodological bases: the construction of a corpus for the detection of deception and credibility assessment¹

Bases metodológicas: la construcción de un corpus para la detección de mentiras y la evaluación de la credibilidad

Pedro Eduardo Hernández Fuentes
Universidad Nacional Autónoma de México –México

Abstract: Meta-analytic approaches reveal that, to identify lies or evaluate the credibility of a testimony, it is more reliable to perform a discursive or verbal material analysis in contrast to one based on non-verbal behavior. Hence, different research has been developed to make scientific contributions in this regard. These efforts make imperative the need to focus on the construction of a linguistic corpus that enables the study of the subject without ignoring the contributions made from cognitive psychology. Therefore, the methodological proposal for the construction of a corpus on the subject will be developed in this text. This is the result of a transdisciplinary work between linguistics and psychology integrated for a total of 54 cognitive interviews using a double-blind procedure.

Resumen: Los acercamientos metaanalíticos revelan que la información verbal es un indicador confiable para identificar mentiras o evaluar la credibilidad de un tes-

¹ Translation from Spanish language by Leon Jacob Ortega Islas.

timonio. De aquí que actualmente se han desarrollado diversas investigaciones para realizar aportaciones científicas al respecto. Estos esfuerzos vuelven imperativa la necesidad de enfocarse en la construcción de un corpus lingüístico que posibilite el estudio del tema sin relegar las aportaciones realizadas desde la psicología (cognitiva). Por ello, en este texto, se desarrollará la propuesta metodológica para la construcción de un corpus en el tema. Ésta es el resultado de un trabajo transdisciplinario entre la lingüística y la psicología que consiste en la realización de 54 entrevistas cognitivas con el método de doble ciego.

1. Introduction

The study of detection of deception and credibility assessment has been of interest to many specialists and has been approached from different disciplines. Although scientific tools have been provided for its study, there is still a widespread false belief that there are key determinants, universal body signals or physiological indicators that are irrefutable proof that an individual is lying. Systematic review to analyze research results quantitatively (meta-analysis) reveals that most of the indicators that researchers typically examine in detection of deception are not related to deception at all (Vrij *et al.*, 2010).

Meta-analytic research also reveal that verbal information is a more reliable indicator to identify deception or assess the credibility of a testimony (DePaulo *et al.*, 2003; Vrij, 2018). Hence, research from forensic linguistics, sociolinguistics, psycholinguistics and mostly, cognitive psychology have currently been developed to make scientific contributions in this regard. These efforts make the need to focus on the construction of a linguistic corpus that allows the study of detection of deception and credibility assessment imperative.

Therefore, this chapter will develop the methodological proposal that allows the creation of a linguistic corpus to identify some characteristic features of the evaluation of truthfulness and lie detection in discourse. This project is the result of an inter- and transdisciplinary work between linguistics and psychology. The project proposal developed at the Language and Cognition Laboratory of the Cognitive Sciences Research Center (UAEM) will be presented, emphasizing the methodology followed for the construction of the sample; an in-depth explanation of the method and general description for the construction of the corpus is presented: type of study, type of participants, data collection procedure and ethical considerations We believe that, before making evaluations on truthfulness or falsehood in discourse, it would be necessary to explore theoretically and meth-

odologically the construction of the discursive corpus in order to begin to explore the still little-known map of deception and truthfulness. Beginning to establish methodological rigor in the construction of this type of samples is not an easy task, although it is necessary for the future experimental or quasi-experimental approach to a subject for which there are many questions and few answers.

We also aspire to introduce to the academic context a subject that has been little addressed in the scientific field, since there have not been enough studies that consider linguistic theory to address this phenomenon: most of the research has been conducted from the perspective of cognitive psychology. There is also a deficiency in the little research conducted on the Spanish language; although some recent proposals consider this language as a field of study there are still few efforts (Hwang *et al.*, 2016; Vrij *et al.*, 2020).

In short, although research has emphasized the preponderance of the analysis of verbal content in contrast to that of nonverbal behavior, there is a gap in this regard and not enough value has been given to the construction of the corpus so that, in the future, the main linguistic indicators that differentiate between a discourse that intends to deceive another and one that does not can be studied. This project will contribute to fill this gap.

2. Detection of Deception and Credibility Assessment

The subject of this paper has a long tradition within the scientific and non-scientific field. The approaches to this subject have been made mainly from philosophy and psychology, although there are also contributions from anthropology, behavioral economics, sociology, and linguistics, to mention a few examples. Possibly, the first major contribution that these works have given is the definition of the terms *lie* and *deception*, which have been used as synonyms, and are understood as an attempt to generate in someone else, from verbal or nonverbal means, a belief that the communicator assumes as false (Vrij, 2008; Masip, 2004) (§3.2). Other contributions concern the answer to questions such as: why do we lie or what are the reasons for lying? (Vrij, 2001, 2008), what are the characteristics of a good liar? (Vrij, 2008), what skills do people have to lie? (Salekin *et al.*, 2008), how often do we lie? (Feldman *et al.*, 2002), what are the basics of lie detection? (Vrij, 2008), and why are we bad lie detectors? (Vrij *et al.*, 2010).

The phenomenon acquired greater visibility from its association with the study of nonverbal behavior, whose most considered channels have been facial expression, physiology, paralanguage and oculusics. Although the study of nonverbal behavior also includes other channels, namely gestures, postures, orientation and movement, proxemics, haptics, and appearance (López *et al.*, 2016), these have been less regarded and studied. Within the study

of facial expression, one of the greatest proponents has been Paul Ekman, who has argued in various publications (Ekman, 2015, 2017; Ekman and Friesen, 1969, 1974; Ekman and O'Sullivan, 1991) that facial expressions of emotions are universal and have a biological, evolutionary, and adaptive origin, as Darwin (1872/2009) stated. Although Darwin's (1872/2009) and, therefore, Ekman's proposals were initially questioned, his findings have now been supported by more than a hundred research studies and different specialists; for example, Reissland *et al.* (2011) conducted a study on facial development based on 4-D ultrasound visualization of fetal facial movements.

These investigations have led to state that one of the most reliable ways to detect deception is the study of *microexpressions*, which are rapid facial movements lasting less than one-fifth of a second, which are important because they convey important information about what a subject is truly feeling or experiencing emotionally and is trying to hide (Ekman, 2017). Based on this, it has been suggested the idea that observable microexpressions on the face are more reliable indications of deception than other channels. This, moreover, is supported by the Filtering Hypothesis, which argues that, when a person lies, he or she experiences emotions that he or she tries to hide because they could reveal the truth; however, these are leaked through the subject's face for a brief moment (Ekman and Friesen, 1969).

However, the analysis of microexpressions as indicators of deception is still under discussion, since deception can generate positive or negative emotions, or even these may not be present and, therefore, the analysis of them is not the best way to determine when a person is hiding the truth (Burgoon, 2018; Vrij *et al.*, 2010). In addition, it remains to delve into the relevance or not of other indicators of nonverbal behavior that have been less studied, such as those that DePaulo *et al.* (2003) registered: the movements of arms, hands, fingers, fingers, legs and feet and the use of illustrators. Therefore, Vrij *et al.* (2010) reviewed which is the most successful way for detection of deception when a subject tries to detect it without the help of technology: nonverbal behavioral analysis or discursive analysis, concluding that a promising way was discursive analysis.

Following the above, Vrij (2018) presented a literature review on the keys in detection of deception and pointed out that the projects that study the differences at the discursive level are the ones that are currently predominant, as there is scientific evidence on their level of reliability. This is also confirmed by the meta-analytical study of DePaulo *et al.* (2003), in which, from the review of the importance of 158 behaviors (verbal and nonverbal), it was concluded that the analysis of the verbal in contrast to the nonverbal is more relevant. So, is detection of deception a problem of linguistics?

2.1 Is detection of deception a problem of linguistics?

The fact that the subject has been widely approached from psychology does not imply that it is not a problem of linguistics. From this area, some research has been carried out, although it is not very abundant, since the study of detection of deception as a linguistic phenomenon has been relegated; hence there is a need to offer more specific contributions from this discipline that give a linguistic description of the phenomenon. It is likely that the limited existence of linguistic studies of lying is the result of the methodological difficulties involved in the design of experiments and the analysis of the information obtained (Infante, 2015). The still low number of contributions made from this area regarding the subject and some peripheral subjects could be listed more and more frequently; however, increasing interest in the construction of a corpus other than English – the language in which the experiments and samples have been mostly designed – may allow us to generate a more assertive approach to the matter.

Among the linguistic contributions, those developed from forensic linguistics stand out, for example, Picornell (2013) has studied the detection of deception in written witness statements and has proposed ways to look for signs of deception from the narrative characteristics of the witnesses. The author has criticized that one of the shortcomings that exist in several of the research studies is that they are conducted with university students because they are the closest participants, although they do not reflect the reality. For this reason, in the present study, the two variables to be controlled are not related to educational level, but to age and sex (§3.3). Also noteworthy are the contributions of Fitzpatrick (2009), who attempted to test the accuracy of some linguistic cues linked to deception.

From a more technological perspective, a number of tools have been developed, for example, the Linguistic Inquiry and Word Count (Pennebaker *et al.*, 2001), used to automate in a simple way the lexical analysis of deceptive text; the Voice Stress Analyzer (NITV Federal Services, 2020), whose hypothesis is that vocal stress indicators reveal deception; and the CSC Deceptive Speech (2013), a corpus developed to distinguish deceptive speech from non-deceptive speech based on machine learning techniques on features extracted from the corpus. These endeavors, which aim to identify and quantify linguistic indicators of deception, have generated several computational programs from different research areas and laboratories in the last fifteen years with the direct or indirect purpose of achieving a better identification of lies: Agent99 Analyzer, General Architecture for Text Engineering (GATE), iSkim or CueCal, Coh-Metrix, Automated Deception Analysis Machine (ADAM) (Hauch *et al.*, 2015).

3. Methodological proposal

The creation of this corpus responds to the interest and the need to create resources that generate research related to truthfulness and deception in discourse, since, as stated in the introduction, most of the current research indicates that the analysis of verbal content can provide more clues in the detection of deception and the evaluation of credibility. Thus, beginning to defragment and study how Spanish speakers lie in quasi-experimental conditions is a timely, though limited, approach for resource generation and future research purposes in this field.

The idea that there is no single totally reliable signal for deception detection is the most useful one because of the very difficulties of lie detection. In this sense, the set of several verbal and non-verbal indicators is the most accurate way to deal with this phenomenon; although the focus of this work is, in principle, linguistic, by obtaining recorded audiovisual material (§3.3), other types of approaches will be possible in the future. It should also be noted that, as mentioned, most of the research reviewed seeks to find patterns that help to determine whether there are indicators of deception, leaving aside the evaluation of truthfulness in discourse. This is also intended to be controlled in the present research.

Thus, the creation of this sample seeks to create a database with a general criterion specific to the Laboratory of Language and Cognition that: 1) favors projects related to the topic; 2) speeds up the necessary methodological processes of a research related to the topic; 3) allows the approach of inter- and transdisciplinary research from the same material whose methodological decisions have a justification; 4) allows finding characteristic patterns of truthful and fallacious discourse of a specific society and with a particular topic. In principle, the scope of the set of texts is limited to the collaborators of the Laboratory, i.e., only members will be able to consult it, since there is no platform on which it can be disseminated. Nevertheless, in the future, a greater transcendence is intended.

The first phase of this research involved the design of the interview and the selection of the participants (§3.3 and §3.4); the second phase involved sending more specific information through the informed consent form (§3.3 and Appendix 2. CI); the third phase involved conducting the cognitive interview divided into two sections (Appendix 3. GE): implementation of the double-blind method and conducting the interview; finally, the fourth phase involved the transcription and basic labeling that will allow for future analyses (Appendix 4. CT).

Upon completion, fifty-four narratives of experience were obtained from twenty-seven people who had some experience of the September 19, 2017, earthquake that occurred in Mexico; each participant provided one truthful narrative (twenty-seven total) and one

fallacious narrative (twenty-seven total). The testimonies were divided into three different groups (Table 2) to be able to perform comparative analyses.

3.1 Cognitive interviewing as a method for eliciting deceptive discourse

The lack of evidence that proves the usefulness of non-verbal parameters in lie detection and credibility assessment has generated the development of research that bets on the use of cognitive strategies. This has led to remarkable differences between those who express (verbally) a truth or a lie (Vrij, 2018) and, therefore, has prompted the design of experiments that assess these distinctions: telling a story backwards rather than in chronological order (Vrij *et al.*, 2012; Vrij *et al.*, 2008), looking at the direction of gaze (Vrij *et al.*, 2010), asking unexpected questions of the participant (Lancaster *et al.*, 2013), asking the subject to perform a secondary task (drawing, for example) during the interview (Lancaster *et al.*, 2013), and providing a greater number of possible details in a story (Leal *et al.*, 2015).

Throughout all of these approaches, the role of the interviewer is critical. For example, one could highlight the difference between the cognitive interview model and the Reid technique model of interviewing and interrogation, which is still used despite its proven ineffectiveness. So, it is important for the interviewer to take an active role and ask questions that generate distinctive reactions between the person who is lying and the person who is telling the truth (Masip and Herrero, 2015). This should be supported by protocols based on solid theoretical models, cognitively based, and supported by research, such as the Activation-Decision-Construction Model (ADCM) proposed by Walczyk and those previously discussed.

It is important to consider the limitations noted about the cognitive models currently developed, since specifying the reasons why lying is cognitively more complex is not the same as elaborating or contrasting models that specify the cognitive processes responsible for the distinctions between lying and telling the truth that clarify answers to questions such as what cognitive processes are activated when a person lies? (Blandón-Gitlin *et al.*, 2017).

As part of the development of research studies that focus on the use of cognitive strategies, we can find the cognitive interview, designed by Geiselman *et al.* (1984) and Fisher and Geiselman (1992) with the purpose of obtaining quality information from the interviewee; in addition to developing an alternative interview method to the existing ones, focused on the mental processes of the witnesses instead of the events that occurred (Fisher and Geiselman, 2019). In its first version, channeled toward criminal investigation, the proposal contained four basic techniques: 1) context reinstatement, 2) telling everything, 3) change of perspective, and 4) change of order. In the second version of the interview (Fisher and Geiselman, 1992), called the enhanced cognitive interview, social and commu-

nificative factors were included, which were intended to improve the social interaction between the interviewer and the interviewee, improve the interviewee's memory and other cognitive processes, and achieve effective communication:

Table 1. Cognitive interview techniques (Fisher & Geiselman, 2019).

| No. | Technique | Description | Improved psychological process |
|-----|---|--|----------------------------------|
| 1 | Rapport | It aims to create a good emotional climate and develop a good relationship between the interviewee and the interviewer. | Social interaction |
| 2 | Active participation of the interviewee | The interviewee actively generates information throughout the interview: he/she does not only answer the interviewer's questions. | Social interaction |
| 3 | Report everything | The interviewee includes all the memories that come to mind, as he/she is asked to report all the facts, whether he/she considers them important or not. | Memory and communication |
| 4 | Reset the context | The interview aims to re-establish the context of the original experience. | Memory |
| 5 | Describe in detail | Seeks a detailed account of events from the interviewee. It can sometimes be initiated from a model statement (Leal, Vrij, Warmelink, Vernham, & Fisher, 2015). | Communication |
| 6 | Close your eyes | The interviewee is asked to close his/her eyes. This instruction should be done after the relationship between the interviewee and the interviewer has been developed. | Cognition |
| 7 | No interruptions | The interviewee should not be interrupted during the interview. | Social interaction and cognition |
| 8 | Do not guess | It is made clear to the respondent that it is okay to say "I don't know" and not to guess the answer. | Cognition |
| 9 | Open questions | It calls for mainly open-ended questions; closed-ended questions will be asked only as a follow-up. | Social interaction and cognition |
| 10 | Multiple recovery | An attempt is made to encourage the interviewee to search through his or her memory more than once. | Memory |
| 11 | Varied recovery | It is intended to encourage the participant to search through his or her memory in different ways. | Memory |
| 12 | Questions compatible with the interviewee | It calls for questions that are compatible with the respondent's current accessibility. | Memory |
| 13 | Avoid suggesting questions | Avoid asking questions that suggest a specific answer. | Memory |
| 14 | Compatible output code | It allows respondents to produce their knowledge in the same form in which it is stored (often non-verbal). | Communication |

Over the years, modifications have been made to the cognitive interview and a consensus has been reached on its effectiveness in contrast to other types of interviews such as structured interviews (Köhnken *et al.*, 1999). It has also been successful in increasing the amount of correct information recalled by the interviewee (Fisher *et al.*, 2011), it has proven to be effective in different contexts and in both criminal and non-criminal investigations (Fisher and Geiselman, 2019). Likewise, it has been widely used in the field of lie detection. Therefore, in this paper, we used this type of interview to obtain the required information.

3.2. Type of study

The type of study of this research is non-probabilistic quasi-experimental in which a corpus was obtained by convenience from the manipulation of two variables of interest: age and sex. For this, in each interview, a pre-post evaluation was conducted from which the baseline of the participants will be obtained according to the evaluation between the narration of the true story (experience of the earthquake of September 19, 2017, in Mexico) and the false version of the same story; both were conducted randomly, that is, in some cases it was decided that the interviewee first lied and then told the truth and vice versa to observe whether this has an effect on the discourse. It should be added that this work does not aim to evaluate the memory or recollection of the participants, but rather their intention to lie or tell the truth; this justifies the decision to use an event that occurred well in advance (see definition of lying, §2). Likewise, the participants were intended to be their own control.

The scientific method used to prevent the results of future research from being influenced by observer bias was the so-called double-blind method: in the collection of the corpus, the participants were unaware of the research topic (Appendix 1. D) while the interviewer and analyst are still unaware of the type of discourse they formulated first, true or false, as the information was determined by an instructor outside the interviewer.

3.3 Participants and interview

A non-probabilistic convenience sampling was carried out. To this end, 27 volunteers (Table 2) were invited to participate using a poster published on social networks, with the following requirements or inclusion criteria: internet access, time availability of approximately one hour, being of one of the requested ages, agreeing to sign an informed consent form (Appendix 2. CI) with the request to videotape their participation for strictly academic purposes, to have a camera and audio in the device to be connected and to have the video call program to conduct the meeting via this means. The exclusion criteria, in addition to non-compliance with any of the above, were neurological problems or language

pathology. Since these were self-declarations, the reliability of this information could not be controlled. Based on these requirements, men and women were selected from each of the three groups shown in Table 2. It should be noted that the initial intention was to obtain 30 volunteers, but only the number indicated was achieved and it was necessary to exclude some of the participants. In the future, we intend to complete the number of participants in order to have a fully gender-balanced sample.

Table 2. Participants.

| Group | Age | Sex | No. of participants |
|-------|-------|-------------------|---------------------|
| 1 | 20-25 | 5 women and 5 men | 10 |
| 2 | 35-40 | 5 women and 5 men | 10 |
| 3 | 50-55 | 5 women and 2 men | 7 |

The project manager determined the eligibility of the participants according to the inclusion and exclusion criteria indicated, based on the answers provided by the volunteer. None of the three groups included vulnerable participants.

A virtual Zoom session was organized for each of the volunteers to conduct the interview. The first face-to-face (virtual) approach was by a person other than the interviewer, known as the “instructor”, to give the participant the instructions developed in the interview guide, the instructor’s guide (Appendix 3. GE). Once his/her participation was completed, the instructor informed the interviewer that he/she had finished so that he/she could enter the session via Zoom and continue with the meeting as detailed in the guide.

The two participant narrations (one true and one false) were both recorded on two different recordings. Each was labeled as follows: CMC000ivA. This label is comprised of basic information to systematize the use of the material, consisting of: 1) the letters CMC refers to the name of the corpus “Corpus mentiras y credibilidad”; 2) the sequence of four numbers corresponds to the number of the video and changes according to the number of testimony; 3) the letter *v* corresponds to the clarification that it is a video; 4) the capital letter corresponds to the letter assigned to each one of the participants.

Once the material was obtained, a Word transcription was made with the corresponding criteria (Appendix 4. CT). These files were labeled CMC000itA, which is the same as the previous label, but with a change in the lowercase letter, which implies that it is a transcription. The transcription process involved two participants: the transcriber and the reviewer.

3.4 Data collection procedure

The participation of the volunteers was videotaped with the Zoom program. The instructor and the interviewer used the interview guide to help them (Appendix 3. GE). The cognitive interview proposals (§3.1) were considered in the elaboration of these materials; they were also reviewed and commented on by three experts.

As for the transcription criteria (Appendix 4. CT), great attention was paid to ensure that the use of marks was the minimum necessary to achieve the purposes of this project, while remaining rigorous. Thus, most of the elements linked to phonetic-phonological characteristics were omitted. Likewise, the participants were given the “Informed Consent” (Appendix 2. CI). All the forms are attached as annexes.

3.5 Ethical considerations

Regarding ethical considerations, this research had minimal risk for the participants, since only documentary research techniques were used (cognitive interview) in which sensitive aspects of behavior were not addressed. The research protocol was sent to the Centro de Investigación Transdisciplinar en Psicología, Universidad Autónoma de Morelos, on September 4, 2020, and was approved on November 30, 2020.

3.6 Current track and future projections

As mentioned at the beginning, this work is mainly of a methodological nature, as it is considered that, since this is a subject that has been little addressed in linguistic and corpus studies, the first approach to follow is to make a proposal that allows us to obtain the truthful and fallacious discourse. In spite of this, some of the results obtained have to do with the type of words present in the total narration, the number of total words, the lexical variety, the approximate duration of the narration and the number of words per minute (Tables 3 and 4).

Table 3. Group 1. Women aged 20 to 25 years.

| | Type | Token | Lexical variety | Approximate duration | Words per minute |
|-----------|------|-------|-----------------|----------------------|------------------|
| CMC0004-B | 448 | 1826 | 4.07 | 11 | 166 |
| CMC0007-D | 303 | 1032 | 3.40 | 9 | 114.66 |
| CMC0009-E | 393 | 1452 | 3.69 | 8 | 181.50 |
| CMC0014-G | 715 | 3803 | 5.31 | 22 | 172.86 |
| CMC0017-I | 479 | 1947 | 4.06 | 11 | 177 |

Table 4. Group 2. Women aged 20 to 25 years.

| | Type | Token | Lexical variety | Approximate duration | Words per minute |
|-----------|------|-------|-----------------|----------------------|------------------|
| CMC0003-B | 385 | 1526 | 3.96 | 10 | 152.60 |
| CMC0008-D | 533 | 2515 | 4.71 | 19 | 132.36 |
| CMC0010-E | 387 | 1255 | 3.24 | 7 | 179.28 |
| CMC0013-G | 525 | 2393 | 4.55 | 14 | 170.92 |
| CMC0018-I | 466 | 2072 | 4.44 | 10 | 207.20 |

The above tables show that we started from a general approach to proceed to a particular one in which potential linguistic indexes are codified to establish their quality. Some of them are part of the psychological, criminological and, to a lesser extent, linguistic literature that have been constantly mentioned and are currently considered as warning flags: full pauses, negation, adverbs, verb tenses, pronouns, number of syllables, number of sentences, number of big words, number of syllables per word, number of short sentences, number of long sentences, average number of words per sentence, conjunctions, simple sentences and adjectives (Burgoon *et al.*, 2003; Fitzpatrick and Bachenko, 2009; Picornell, 2013; Villar and Castillo, 2016). Currently, the coding of filled pauses, pronouns, adverbs, reported memory, in addition to those previously mentioned, is part of the tasks of the coordinator of this research and the first results are expected to be available in March 2022.

4. Conclusions and discussion

Although research has emphasized the preponderance of the analysis of verbal content in contrast to that of nonverbal behavior, there is a lack of studies that delve into the considerations of linguistic theory and that focus, as a first step, on the construction of a corpus that allows the study of the main linguistic indicators that distinguish between a discourse that intends to deceive and one that does not.

It is necessary to consider that detection of deception is complex, so it could be easy to fall into the Othello error, a concept coined by Ekman (2015) to refer to the errors in which the evaluator may fall if he/she does not consider that a person who is telling the truth may “appear” to be a liar when only one level of analysis is considered. In this sense, it is easy for biases such as gaze direction or the different comfortable certainties mentioned in this work to induce error. Hence, this paper seeks to reduce this type of errors through a promising approach, which is a verbal and cognitive one in which the analyst’s bias (with the double-blind method) is reduced.

Regarding detection of deception and truthfulness assessment, it is clear that there are currently different research studies that question the analysis of *microexpressions* or *para-linguistics* as viable channels of analysis. Although this paper does not go into this issue in depth, it is considered that the study of these channels can always provide valuable information if they are considered as part of a whole. This implies aiming at a constellation of evidence in which the analysis of linguistic behavior is as important as the analysis of non-linguistic behavior: an isolated analysis of non-verbal behavior would be just as dangerous as concentrating exclusively on a strictly linguistic analysis. For methodological reasons, however, in this work greater weight has been given to the construction of the corpus from a more linguistic angle, without disregarding the other channels. Hence the audiovisual recording of the participants who took part in this project.

In sum, this work makes different contributions. First, it establishes methodological rigor in the construction of a corpus for the identification of linguistic strategies linked to lies and truthfulness. This implied the careful selection of participants, the elaboration of instruments such as the interview guide with a solid theoretical basis, and the submission of the project itself to an ethics committee.

Moreover, by obtaining two types of discourse (one truthful and one false), it is intended that soon it will be possible to study both differences in the same subject, that is, to know the linguistic baseline of the participant when he/she tells the truth in order to recognize the relevant and significant differences when the same subject lies. It should be considered that, in the future, specialists in “detection of deception” should focus on assessing truthfulness in discourse rather than on identifying lies. This project thus emphasizes both fallacious and truthful discourse.

Finally, an advantage of the project is the transdisciplinary ethos that it aims to have so that, over time, more collaborative work with different disciplines can be carried out to understand a phenomenon that is present in our daily interactions.

Appendix

Below is a summary of each of the appendices attached to the research in Spanish.

1. D. Diffusion

This appendix corresponds to the poster used for the search of volunteers. It indicates the requirements, includes contact information and general information about the research. The poster was circulated by the Language and Cognition Laboratory of the Center for Research in Cognitive Sciences (UAEM).

2. **CI. Informed Consent Form**

This appendix contains the informed consent form. This appendix includes the consent of the volunteers to participate in the collection of interviews as part of the Language and Cognition Laboratory project. It specifies the risks, type of research technique, rights, benefits, and general structure of the interview.

3. **GE. Interview Guide**

This appendix is divided into two sections: Instructor's Guide and Interviewer's Guide. In the first section, the general instructions to be given by the instructor to the interviewer are detailed, that is, to welcome him/her and the instruction to lie or tell the truth in each of the narratives according to the order decided by the instructor himself/herself. In the second section, more specific information about the project is mentioned, the instruction given by the instructor is reinforced without discovering the double-blind, a model description of the type of narrative expected is made, the participant's acceptance is asked again, and the interviewee's narrative begins with the completion of the question in which the narrative of the experience of the earthquake of September 19, 2017, is requested.

4. **CT. Transcription Conventions**

This section details the transcription conventions used. The criteria used in terms of spelling and punctuation, phonic and lexical labeling, labeling of discursive dynamics, and format criteria are mentioned.

— *References*

- Burgoon, J. K. (2018). Microexpressions Are Not the Best Way to Catch a Liar. *Frontiers in Psychology*, 9, 1-5.
- Blandón-Gitlin, I., López, R. M., Masip, J. y Fenn, E. (2017). Cognición, emoción y mentira: implicaciones para detectar el engaño. *Anuario de Psicología Jurídica*, 27(1), 95-106.
- Columbia University, SRI International, and University of Colorado Boulder. (2013). *CSC Deceptive Speech LDC2013S09*. Recurso electrónico. Linguistic Data Consortium. <https://doi.org/10.35111/q500-9a28>
- Darwin, C. (1872/2009). *La expresión de las emociones*. Laetoli.
- DePaulo, B., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to Deception. *Psychological Bulletin*, 129(1), 74-118.
- Ekman, P. & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32, 88-106.
- Ekman, P. & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, 29(3), 288-298.

- Ekman, P., & O'Sullivan M. (1991). Facial expression: methods, means, and moes. In R. S. Feldman, & B. Rimé, (Eds.), *Fundamentals of Nonverbal Behavior* (pp. 163-199). Cambridge University Press.
- Ekman, P. (2015). *Cómo detectar mentiras. Una guía para utilizar en el trabajo, la política y la pareja*. Paidós.
- Ekman, P. (2017). *El rostro de las emociones. Qué nos revelas las expresiones faciales*. RBA.
- Feldman, R. S., Forrest, J. A., & Happ, B. R. (2002). Self-presentation and verbal deception: Do self-presenters lie more? *Basic and Applied Social Psychology*, 24(2), 163-170.
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Charles C. Thomas.
- Fisher, R. P., & Geiselman, R. E. (2019). Expanding the Cognitive Interview to Non-Criminal Investigations. In J. Dickinson, N. Schreiber Compo, R. Carol, B. L. Schwartz, & M. McCauley (Eds.), *Evidence-based Investigative Interviewing Applying Cognitive Principles* (pp. 1-28). Routledge, Taylor & Francis Group.
- Fisher, R. P., Milne, R., y Bull, R. (2011). Interviewing cooperative witnesses. *Current Directions in Psychological Science*, 20, 16-19.
- Fitzpatrick, E. & Bachenko, J. (2009). Building a forensic corpus to test language-based indicators of deception. In: S. T. Gries, S. Wulff & M. Davies (Eds.), *Corpus-linguistic applications. Current studies, new directions* (pp. 183-196). Rodopi.
- Geiselman, R. E., Fisher, R. P., Firstenberg, I., Hutton, L., Sullivan, S. J., Avetissian, I. V., & Prosk, A. L. (1984). Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview. *Journal of Police and Science Administration*, 12, 74-80.
- Hauch, V., Sporer, S. L., Michael, S. W. & Meissner, C. A. (2014). Does training improve detection of deception? A meta-analysis. *Communication Research*, 43(3), 283-343.
- Hwang, H. C., Matsumoto, D. & Sandoval, V. (2016). Linguistic Cues of Deception Across Multiple Language Groups in a Mock Crime Context. *Journal of Investigative Psychology and Offender Profiling*, 13, 56-69.
- Infante Arriagada, P. (2015). La mentira como fenómeno lingüístico: algunos aspectos centrales para su descripción. *LL Journal*, 10(2), 1-20.
- Köhnken, G., Milne, R. Memon, A., & Bull, R. (1999). A meta-analysis on the effects of the Cognitive Interview. *Psychology, Crime, & Law*, 5, 3-27.
- Lancaster, G. L., Vrij, A., Hope, L. & Waller, B. (2013). Sorting the liars from the truth-tellers: The benefits of asking unanticipated questions on lie detection. *Applied Cognitive Psychology*, 27, 107-114.
- Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & Fisher, R. P. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology*, 20(1), 129-146.
- López Pérez, R. M., F. Gordillo León y M. Gau Olivares (coords.). (2016). *Comportamiento no verbal. Más allá de la comunicación y el lenguaje*. Pirámide.
- Masip, J., Garrido, E. y Herrero, C. (2004). La detección de la mentira mediante la medida de la tensión en la voz: una revisión crítica. *Estudios de Psicología*, 25(1), 13-30.
- Masip, J., y Herrero, C. (2015). Nuevas aproximaciones en detección de mentiras I. Antecedentes y marco teórico. *Papeles del Psicólogo*, 36(2), 83-95.
- NITV Federal Services (2020). *Voice Stress Analyzer*. <https://www.cvsai.com/>.
- Pennebaker, J. W., Francis, M. E. & Booth, R. J. (2001). *Linguistic Inquiry and Word Count*. Lawrence Erlbaum Associates.

- Picornell, I. (2013). Analysing Deception in Written Witness Statements. *Linguistic Evidence in Security, Law and Intelligence*, 1(1), 41-50.
- Reissland, N., Francis, B., Mason, J. & Lincoln, K. (2011). Do Facial Expressions Develop before Birth? *PlosOne*, 6(8), 1-7.
- Salekin, R. T., Kubak, F. A. & Lee, Z. (2008). Deception in children and adolescents. In R. Rogers, & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (p. 343-364). The Guilford Press.
- Villar, G., & Castillo, P. (2016). The Presence of 'Um' as a Marker of Truthfulness in the Speech of TV Personalities. *Psychiatry, psychology, and law: an interdisciplinary journal of the Australian and New Zealand Association of Psychiatry, Psychology and Law*, 24(4), 549-560.
- Vrij, A. (2001). Detecting the liars. *Psychologist*, 14, 596-598.
- Vrij, A. (2008). *Wiley series in the psychology of crime, policing and law. Detecting lies and deceit: Pitfalls and opportunities* (2^a ed.). John Wiley & Sons Ltd.
- Vrij, A. (2018). Deception and truth detection when analyzing nonverbal and verbal cues. *Applied Cognitive Psychology*, 33(2), 160-167.
- Vrij, A., Fisher, R. P., Mann, S., Deeb, H., Jo, E., Castro Campos, C., & Hamzeh, S. (2020). The Efficacy of Using Countermeasures in a Model Statement Interview. *The European Journal of Psychology Applied to Legal Context*, 12(1), 23-34.
- Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and Opportunities in Nonverbal and Verbal Lie Detection. *Psychological Science in the Public Interest*, 11(3), 89-121.
- Vrij, A., Leal, S., Mann, S. A. y Fisher, R. P. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse order technique naturally. *Psychology, Crime & Law*, 18, 579-594.
- Vrij, A., Mann, S. A., Leal, S. & Fisher, R. P. (2010). "Look into my eyes": Can an instruction to maintain eye contact facilitate lie detection? *Psychology, Crime & Law*, 16, 327-348.