

Cross-Language Applicability of Linguistic Features Associated with Veracity and Deception

David Matsumoto, Hyisung C. Hwang & Vincent A. Sandoval

Journal of Police and Criminal Psychology

The Official Journal of the Society for Police and Criminal Psychology

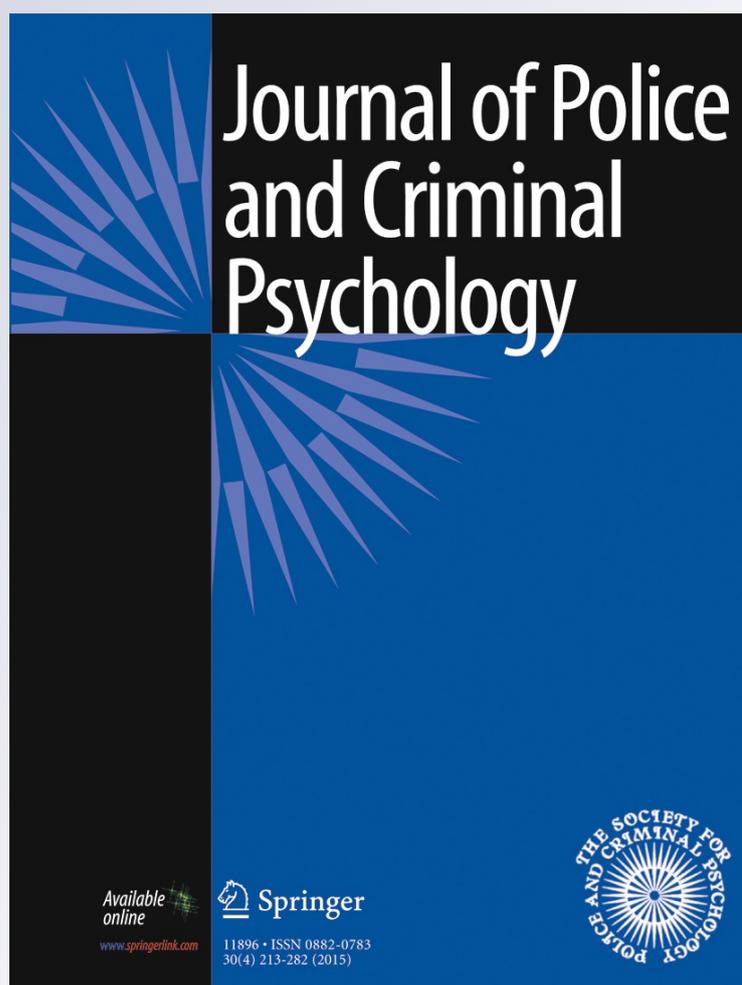
ISSN 0882-0783

Volume 30

Number 4

J Police Crim Psych (2015) 30:229-241

DOI 10.1007/s11896-014-9155-0



Your article is protected by copyright and all rights are held exclusively by Society for Police and criminal Psychology. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Cross-Language Applicability of Linguistic Features Associated with Veracity and Deception

David Matsumoto · Hyisung C. Hwang ·
Vincent A. Sandoval

Published online: 28 August 2014
© Society for Police and criminal Psychology 2014

Abstract One technique for examining written statements or interview transcripts for verbal cues of veracity and lying involves the analysis of linguistic features and grammatical structures associated with word usage. This technique is commonly referred to as Statement Analysis (SA). There are varying degrees of empirical support for different SA techniques and for specific linguistic markers; what is less known in the literature is the degree to which verbal indicators of veracity and lying vary across languages. We examined this research question. Participants from three language groups – English, Spanish, and Chinese – witnessed a video portraying an actual crime and then wrote false and true statements about what they had witnessed in their respective languages. The statements were coded using various linguistic features of SA. The selected linguistic features discriminated between true and false witness statements and the effect sizes were relatively large. Importantly, language did *not* moderate the relationship between veracity and the coded features, indicating cross-language similarity in the efficacy of SA features to differentiate truths from lies.

Keywords Deception · Culture · Language · Statement analysis

One technique for examining language for clues to lying involves the analysis of linguistic features and grammatical

structures associated with word usage, commonly referred to as Statement Analysis (SA). SA is based on the premise that word use and grammar structures differ when people lie as opposed to when they tell the truth. Because words make up sentences and sentence construction follows a predetermined set of grammatical rules, a careful examination of word use and grammar structures should identify specific features that can help detect deception.

SA has its roots in psycholinguistic research in the early 1900s but has received more contemporary reception within forensic psychological and law enforcement circles as a result of the work of Undeutsch (1989) and a technique known as Statement Validity Analysis (SVA). SVA was founded on a hypothesis that statements based on actual memories differ from fabricated or fantasized statements (Undeutsch, 1989). The crucial parts of SVA involve a criteria-based content analysis (CBCA) and an evaluation of CBCA outcomes using a Validity Check-List with criteria organized around categories such as general, unusual, motivational and stylistic features. In addition to SVA a number of other techniques that involve the analysis of the grammatical structures to make inferences about deception and truthfulness have emerged, including Theoretical Verbal Analysis (TVA; Connelly, et al., 2006), Reality Monitoring (RM; Johnson & Raye, 1998), Scientific Content Analysis (SCAN; Sapir, 1996), and Investigative Discourse Analysis (IDA), which is an extension of CBCA and similar to SCAN (Rabon, 1994).

Research examining the efficacy of various SA techniques has provided evidence for many of them to detect truths from lies at better than chance accuracies (Porter & Yuille, 1996; Vrij, 2007; Vrij & Mann, 2006; Zaparniuk, Yuille, & Taylor, 1995). For example, CBCA has been linked to empirically based knowledge about naturalistic memory and to a fair amount of research demonstrating the validity of many of its

D. Matsumoto (✉) · H. C. Hwang
Department of Psychology, San Francisco State University, 1600
Holloway Avenue, San Francisco, CA 94132, USA
e-mail: dm@sfsu.edu

D. Matsumoto · H. C. Hwang
Humintell, LLC, Berkeley, CA, USA

V. A. Sandoval
Federal Bureau of Investigation (retired), Washington, DC, USA

criteria (Porter, Birt, Yuille, & Lehman, 2000; Porter & Yuille, 1996; Zaparniuk, et al., 1995). RM is also based on a solid empirical base of knowledge about memory (Johnson, 1988; Johnson & Raye, 1981) and reviews examining its usefulness in detecting deception have confirmed the validity of many of its criteria (Masip, Sporer, Garrido, & Herrero, 2005; Sporer, 2004).

What is less known in the literature is whether verbal indicators of veracity and lying vary across languages, and whether or not the same SA features can be used to differentiate truths from lies across languages, because most of the research to date has analyzed source materials produced in one language (usually English by native English speakers). To be sure there are studies examining the verbal indicators of veracity and lying in non-English languages (Masip, Bethencourt, Lucas, Sanchez-San Segundo, & Herrero, 2012; Ruby & Brigham, 1997; Schelleman-Offermans & Merckelbach, 2010). For instance, Masip, et al. (2012) asked Spanish students to write a truthful or deceptive story; subsequent analyses demonstrated that truthful and false stories differed on plausibility, details, consistency, emotions, and structure. Schelleman-Offermans and Merckelbach (2010) asked students in The Netherlands to write a true or false story (in Dutch) about an aversive situation in which they had been victims (e.g., of gossip, bullying, robbery, etc.). The statements were coded using nine CBCA criteria; three differentiated true statements from false: logical structure, contextual embedding, and attribution of the perpetrator's state. Although these studies are suggestive of the potential cross-language applicability of SA, however, comparing results across these studies to make generalizations is problematic because study differences confound the languages examined and their results (but study differences also speak to the robustness of the findings).

We remedied this situation by analyzing the linguistic indicators of veracity and lying in a realistic, moderately high stakes scenario and by examining three very different languages within the same study. There are important theoretical reasons to investigate this issue. Cross-language consistency may provide evidence for potential pancultural similarity in the underlying psychological effects of lying and similarity as reflected in the linguistic choices that mark those effects. Such effects would suggest cross-cultural similarity in the structure of memory, the recall of information from memory, and the psychological demands placed on individuals who lie about that recall, and would point to a potential universal mechanism of lying that can be identified by specific linguistic markers. If the rules of grammar and deep structure of language (Chomsky, 1957, 1972) and the principles of memory and recall (Undeutsch, 1989) are similar across cultures, then verbal indicators of truths and lies may occur regardless of culture, ethnicity, and language.

In this study individuals from three language groups – English, Spanish, and Chinese – witnessed a crime and were asked to write false and true statements about what they had witnessed. Both statements were written in the native language of the participants. The three language groups were chosen to sample a broad range of cultural/linguistic differences that may influence indicators of veracity and deception. These languages also represented major language/cultural groups around the world, as well as the U.S.

Linguistic and Grammatical Markers of Veracity and Lying Used in this Study

There are some commonalities among the various SA techniques as they are based on a relatively common understanding of the nature of human memory and verbal recall of that knowledge. Differences among these systems occur concerning the specific linguistic categories considered indicative of veracity or lying and in the amount of scientific evidence that exists for all the various features of each system, especially across different languages and cultural/ethnic groups. Because SA techniques allow for the analysis of many different types of linguistic and grammatical markers with commonalities, and because of differences across studies in the degree of empirical support for specific categories within specific techniques, we selected for use in this study an eclectic group of SA categories from different techniques deemed most relevant for the source materials produced.

More specifically, for this study, we were not concerned with testing the applicability of any one SA technique across languages, but rather whether or not a core set of SA categories that exist across techniques was applicable across languages. Thus the categories we selected for use in this study were those that occurred across different SA techniques (e.g., CBCA and Reality Monitoring) and for which there was empirical support for use with people of different cultural/ethnic backgrounds. And, we were interested in including SA categories that have proven to be operationally relevant in the field (i.e., categories that have also been the most effective in actual investigations based on the experiences of the third author). The following categories were selected on an a priori basis for use in this study before any statements were coded.

Indicators of Veracity

Structural Balance In response to an open-ended question designed to elicit information, a writer will typically include information about what transpired prior to (prologue), during (incident), and after (epilogue) the incident (Johnson, 1988). The *prologue* provides contextual information relative to the setting such as details pertaining to time of day, place and the

people involved. The *incident* pertains to that portion of the statement where the actual criminal event takes place and begins at that point in the narrative where an investigator would conclude a crime is taking place and warrant initiating an investigation (Rabon, 1994). The *epilogue* consists of subordinate information such as the writer's emotional reaction to the incident or efforts to contact law enforcement. Research and experience have demonstrated that a good indicator of veracity is balance within the structure of a written statement (Adams & Jarvis, 2006; Rabon, 1994), because there is an expectation that when writers discuss a specific event they will dedicate most of their statement to that event. Balance is determined by ascertaining how much space the writer dedicates to each of the three component parts of a statement. Researchers differ on the precise percentages that each component part should possess but all agree that the incident should be at least equal to or greater than the prologue and epilogue (Rabon, 1994; Sapir, 1996). When a statement contains an inordinately long prologue, that statement will often be a deceptive statement (Adams & Jarvis, 2006; Rabon, 1994; Rudacille, 1994; Sapir, 1996).

Word Count A number of studies have demonstrated that liars use fewer words than truth-tellers (DePaulo, et al., 2003; Newman, Pennebaker, Berry, & Richards, 2003). This is likely the result of liars using a strategy of simply omitting important details from their written statement.

Unique Sensory Detail and Spatial Detail (USD-SD) Researchers have postulated that there are identifiable differences between truthful and fabricated statements by identifying the presence of and location within the statement of specific types of details within those statements (Johnson, 1988; Johnson & Raye, 1981; Porter & Yuille, 1996; Undeutsch, 1989). Unique sensory detail (USD) pertains to specific descriptions generated by the five sensory perceptions (sight, sound, touch, smell, taste and touch). Spatial detail (SD) pertains to specific locations and the physical relationships of objects, people, etc., in relation to one another (Adams & Jarvis, 2006). The expectation is that truthful writers who discuss a specific event will provide requisite detail about that event. While evidence for these kinds of details to differentiate truths from lies comes from several theoretical frameworks, both CBCA (Porter & Yuille, 1996; Undeutsch, 1989; Vrij, 2007) and the Reality Monitoring frameworks (Johnson, 1988; Johnson & Raye, 1981) in particular have provided strong evidence to suggest that individuals who recall previously encoded events truthfully report more sensory and spatial details because these details are encoded in memory along with the factual content of the event.

Emotion The presence of the writer's emotion or affective responses in reaction to the incident, such as fear, anger, shock, or embarrassment can also differentiate truths from lies. Both CBCA (Porter & Yuille, 1996; Undeutsch, 1989; Vrij, 2007) and the Reality Monitoring frameworks (Johnson, 1988; Johnson & Raye, 1981) provide evidence to indicate that individuals who recall previously encoded events truthfully also report more emotions in relation to the event because emotions are encoded in memory when events occur. Other studies have suggested that this is especially true when mentions of emotion are found in the epilogue of the written statement (Adams & Jarvis, 2006).

Indicators of Lying

Extraneous Information A number of studies (DePaulo, et al., 2003; Matsumoto, Hwang, & Sandoval, 2013; Vrij, 2007) have demonstrated that truth tellers provide more details relevant to the question raised, whereas liars provide more information that is irrelevant, which we refer to as extraneous information. Extraneous information is information that does not answer the question posed, and may be used to justify the liars' actions, deflect the question because they may not want to respond to that specific question, help liars distance themselves from the act of lying or the content of the lie, or aid liars in exerting control over the interview (Adams, 1996). Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more extraneous information when writing statements in English.

Equivocation Equivocation refers to information that is not relevant to the question that was posed to elicit the statement. Equivocation words qualify statements, allowing liars to distance themselves from the act or content of lying by tempering the action about to be described or by discounting the message even before it is transmitted (Weintraub, 1989). Equivocation consists of words or phrases such as "maybe", "believe", "kind of", "sort of", "about", or "to the best of my knowledge", which suggest that the interviewee is being intentionally vague or ambiguous. Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more equivocation when writing statements in English.

Non-prompted Negation (NPN) When responding to a question such as, "Tell me what you did in the file room," the expectation is that individuals will respond by providing information pertaining to what they actually did (Rudacille, 1994; Sapir, 1996; Weiner & Mehrabian, 1968). Therefore a response about what the individual did not do (e.g., "I did not see a car hit anyone") does not answer the question and is an example of NPN. Negation in discourse or statements may be an indicator of deception inasmuch as respondents may use it to carefully omit their involvement in a crime (Adams &

Jarvis, 2006), and there are generally more negative statements in deceptive oral narratives than in truthful oral accounts (Hauch, Blandon-Gitlin, Masip, & Sporer, 2012; Newman, et al., 2003; Porter, et al., 2000). Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more NPN both when writing statements in English and in oral interviews.

Moderating Adverbs We identify three types of moderating adverbs. (1) *Intensifying adverbs* such as “very,” “really,” “truthfully,” or “honestly” are typically used when a communicator is attempting to convince another person of something. (2) *Minimizing adverbs* such as “only,” “just,” “simply,” “merely” are typically used to downplay or minimize the role of the actor, who is generally the communicator him or herself. (3) *Editing adverbs* such as “after,” “then,” “next,” “while,” “so,” “thereafter,” or “when” may indicate a temporal lacunae (Rabon, 1994; Schafer, 2007) suggesting that the communicator is intentionally editing information and as such, something that might be crucial to an inquiry may be missing from the discourse. Because lies of omission are more common than lies of commission, and because liars tend to use fewer words than truth tellers (DePaulo, et al., 2003), editing adverbs provide liars with a simple yet strategic means of telling the truth up to a certain point, omitting crucial information and then picking up again by telling the truth. Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more moderating adverbs both when writing statements and in oral interviews.

Passive Voice When describing their actions, people will generally assume responsibility for those actions by employing the active voice (i.e., the agent engaging in the action described by a verb is the subject of the sentence). For example, liars who attempt to conceal their identity as an actor, such as firing a pistol, may attempt to distance themselves from the action by employing the passive voice, “the pistol was fired.” Passive voice occurs when the object of an action verb appears as the subject of the sentence. It may be used when liars attempt to conceal their identity as an actor, distancing themselves from the action of the verb (Connelly, et al., 2006; Rudacille, 1994).

Overview of the Study

Participants from three language groups – English, Spanish, and Chinese – witnessed a crime and were asked to write false and true statements about what they had witnessed, in their native language. Participants were led to believe that their statements would be read by investigators who would make a determination about the believability of the statements, and that there were rewards and punishments for the participants

depending on those determinations. The statements they produced were coded for the linguistic features of veracity and lying described above. We hypothesized that the coded features could differentiate true and false statements across the three languages.

Methods

Stimuli

Initial Pool Because the study involved participants writing about a crime they had witnessed, it was necessary to first conduct a pilot study to select a crime video that could be used. We wanted to use a video that depicted an actual crime, that aroused emotions in the part of witnesses/viewers (as would most crimes), and that allowed for a fair test of the structural features of true and false statements (vis-à-vis structural balance). Thus we conducted a pilot study in order to identify the video that would be used in the main study.

First, we searched the Internet for open source videos of actual crimes in different cultures. Surprisingly we found many such videos, many of which were posted by local police departments requesting the aid of the public in identifying persons of interest in the videos. Different types of crimes were represented including animal cruelty, armed robbery, arson, assault and battery, ATM theft, auto theft, burglary, hit and run, kidnapping, mugging, murder, police brutality, shoplifting, pick pocketing, and vandalism. Our search resulted in obtaining an initial pool of 371 videos.

We then excluded videos that included any language in the video – either audio or written (subtitles) – because such commentary may have biased observers’ reactions. We also excluded videos that were part of news reports (moderated by a newscaster) or that had technical difficulties (e.g., extremely low resolution). This resulted in a smaller pool of 94 videos from the U.S. or England, 48 from China, 6 from the Middle East, and 10 from Central or South Asia.

Although all videos were identified as “crime videos,” in many cases it was not clear that a crime had been committed unless the viewer had background information about the action in the video. For example, a video of an “auto theft” of a person unlocking a car and driving off is innocuous unless the viewer knows that the driver is not the owner of the car. Because it was important to use videos that were clear that a crime was committed just by the observation of the video and not requiring any such background information or assumptions, two coders coded whether a crime had clearly been committed on each of the videos using a 5-point scale labeled 1, not clear at all, to 5, very clear.

Additionally we wanted to use videos that were relatively balanced in the amount of time devoted to the portrayal of the incident and before (prologue) and after (epilogue). As

described earlier, Structural Balance is one of the features that may differentiate true and false statements; thus we needed videos that were balanced themselves so as not to skew balance in the resulting statements produced. In the videos, an “incident” was defined as the act or event when the individual’s behavior in that situation deviated from the norm. Thus we also had coders log the time from the start of the video that the incident occurred and when the incident ended. We then calculated the amount of video times dedicated to the prologue, incident, and epilogue.

Videos were selected for use in the study if the video had a crime rating of 5 from both coders and the percentage of the video dedicated to the prologue and incident was each at least 30% of the entire length of each video. This resulted in the final selection of seven potential videos (country of origin of the video in parentheses):

- Video 1: Guy breaks into a car (China)
- Video 2: A woman shoplifts in a beauty supply store (U.S.)
- Video 3: A woman gets caught stealing from a store (U.S.)
- Video 4: Bangalore hit and run accident on the highway (India)
- Video 5: Guy throws brick into a car (England)
- Video 6: Burger King robbery at gunpoint (U.S.)
- Video 7: Animal cruelty – dog gets beaten to death (China)

We also selected one video to use as practice (motorcycle theft) for observers prior to their observing and rating the seven target videos. Thus eight videos were rated.

Observers, Judgment Tasks, and Procedures A total of 555 observers from the U.S. ($n = 63$), India ($n = 143$), Ecuador ($n = 34$), Mexico ($n = 44$), Bolivia ($n = 30$), China ($n = 209$), and South Korea ($n = 32$) participated. They all self-reported as being born and raised in their respective country and their first language corresponded to the language of their country. Local assistants recruited all observers from Ecuador, Mexico, Bolivia, China and South Korea in country; the U.S. Americans participated in our laboratory in Berkeley, California. The Indians were recruited using Amazon Mechanical Turk.

All survey materials were presented online and participants were provided the following instructions:

“The information gathered will be used for research examining cultural differences in perceptions of criminal acts. You will view several video scenes of acts, such as shoplifting, theft, etc. After each video, you will be asked some very basic questions about your thoughts about what you saw, such as ratings of believability,

realism, probability of actual occurrence in your culture, the meaning of the act and its perceived legality, whether you have actually witnessed such an act in the past or heard about an actual event. You will also be asked basic demographic questions such as age, ethnicity and language. You will NOT be asked your name anywhere.”

After providing implied consent, participants were then shown the practice video. They were told to click the play button when ready, that they can enlarge to full screen by clicking the box [] at the bottom right of the video box, and to click ESC to return when done viewing. After the video played, they were asked to rate how the video made them feel by indicating the extent to which they were currently experiencing any or all of the following emotions on a scale labeled 0, did NOT feel ANY of that emotion, to 8, an extreme amount of that emotion: Guilt, Fear, Anger, Embarrassment, Worry, Contempt, Excitement, Disgust, Amusement, Nervousness, Surprise, Interest, Sadness, and Pride. They also completed a set of attitude and belief ratings not germane to this study.

After completing the ratings, observers were shown the actual videos used in the study and given the same instructions as above for the practice video. The videos were shown in the order listed above, from Video 1 through Video 7, because we considered them to be ordered in terms of emotional intensity, from least to most. We ordered them in this fashion to minimize the impact of emotional videos influencing the ratings of subsequent videos.

After the completion of the ratings for all videos, participants provided some basic demographic information. Completion of their demographics marked the end of their participation in the study.

Final Selection The goal of the analyses was to determine cultural differences or similarities in the ratings of the videos in order to select videos for use in the main study that were relatively cross-culturally invariant and elicited the greatest emotion. To elucidate this issue, we computed intraclass correlations (ICCs) across means of the 14 emotion ratings considering the seven countries as raters, separately for each video. These analyses allowed us to determine if the relative ranking among the 14 emotions was consistent across countries or not. ICCs can be computed in two ways, one using absolute agreement as a basis and a second way using consistency as a basis. We were particularly interested in ICCs based on absolute agreement, as these would indicate the degree to which the relative rankings of the 14 emotion means were similar across cultures *and* anchored to a similar absolute score.

There was considerably high agreement across the countries on the relative rankings across means of the 14 emotions rated for each of the seven videos and across all videos overall

(Table 1). These findings suggested that there was a great deal of consistency across the countries in the means of their emotional profiles for each of the videos.

We then examined the overall marginal means of the 14 emotion ratings for each of the videos to determine which video(s) elicited the greatest overall amount of emotion. (Examination of the marginal means was justified given the high ICC values obtained in the results above.) The videos ranked in size of their emotion marginal means were Video 7 (Animal Cruelty, $M = 4.96$, $SE = .09$), Video 4 (Hit and Run, $M = 4.45$, $SE = .08$), Video 6 (Robbery, $M = 4.26$, $SE = .09$), Video 5 (Brick, $M = 4.07$, $SE = .09$), Video 2 (Shoplifting, $M = 3.84$, $SE = .08$), Video 1 (Car, $M = 3.83$, $SE = .08$), and Video 3 (Stealing, $M = 3.82$, $SE = .08$). Because we wanted to use the video that elicited the highest overall emotion ratings, we initially attempted to obtain IRB approval for use of Video 7 (Animal Cruelty). Unfortunately approval was not obtained for that video; thus Video 4 (Hit and Run) was used in this study.

Participants

The participants in the main study included 43 Chinese (20 females, 22 males, 1 undeclared; mean age = 31.49), 38 English (17 females, 17 males, 4 undeclared; mean age = 41.00), and 42 Spanish writers (25 females, 16 males, 1 undeclared; mean age = 32.45). The Chinese participants were all born and raised in Hong Kong, Taiwan, or Mainland China and reported Chinese as their native language; the English participants were all born and raised in the U.S. and reported English as their native language; and the Spanish participants were all born and raised in Central or South America, and reported Spanish as their native language. All reported being fluent in reading, writing, and speaking in their target language.

Participants were recruited from the local communities in the San Francisco Bay Area through online and hardcopy ads seeking individuals who were 18 years of age or older, and born and raised in a country for which the target language was the official language. The ads stipulated that the study requires

reading and writing in the target language, and that all participants must read and write the target language fluently. Ads appeared in both English and in the target languages. Prior to participation all potential participants were telephone screened according to the inclusion criteria recruited for, and answered the same questions in a standard demographics form obtained as part of the pre-session measures, including self-reported reading, writing, and speaking proficiencies (see below).

Age differed significantly among the three groups, $F(2, 118) = 7.10$, $p = .001$, $\eta_p^2 = .11$. To examine if age covaried with the dependent variables tested, we computed correlations between age and the coded statement analysis categories (described below), separately for the true and false statements and each of the groups. Of the 66 effects (11 coded categories \times 2 statements \times 3 language groups), only 3 were significant (2 for the Chinese, 1 for English, and 0 for Spanish). Thus we concluded that age differences did not covary with the differences in usage of the statement analysis categories.

Measures

At the beginning of the experiment all participants completed the following instruments:

- A basic demographics questionnaire that confirmed ethnic group identity, places of birth and upbringing of themselves and parents, and first and other languages with self-ratings of language proficiency (excellent, good, fair, poor separately for reading, writing, and speaking)
- The General Ethnicity Questionnaire (GEQ; Tsai, Ying, & Lee, 2000); see below.
- An emotion checklist (guilt, fear, anger, embarrassment, worry, contempt, excitement, disgust, amusement, nervousness, surprise, and interest) in which participants self-reported their emotional experiences using 9-point scales labeled 0, None, 4, Moderate Amount, and 8, Extremely Strong. Participants also completed this checklist at the end of the experiment.
- The Machiavellianism Scale (Christie, 1970), a 10-item test assessing individual differences in cunningness, duplicity, or interpersonal manipulation. Previous studies have demonstrated its internal reliability and convergent and discriminant validity with other personality measures (Paulhus & Williams, 2002).
- The NEO-Five Factor Inventory (Costa & McCrae, 1992), a 60-item test assessing the five personality traits found to be universal (Costa & McCrae, 1992): Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. Participants respond to each item using a 5-point scale (0 = strongly disagree to 4 = strongly agree); scale scores are computed using a standard formula. There is substantial evidence for the cross-cultural equivalence in the factor structure and within-country

Table 1 ICCs across the 14 Emotion Means using Countries as Raters

Video	ICC for absolute agreement	ICC for consistency
1	0.890	0.930
2	0.902	0.937
3	0.910	0.941
4	0.913	0.948
5	0.906	0.945
6	0.910	0.944
7	0.927	0.957
All	0.913	0.946

validity of the NEO-FFI (McCrae & Costa, 1997; McCrae, et al., 2005).

- The Social Dominance Orientation Scale (Pratto, Sidanius, Stallworth, & Malle, 1994), a 16-item test that measures individual differences in preferences for hierarchies within social groups and dominance of lower-status groups. After reverse coding specific items, all items are summed to produce a score. There is ample evidence for the internal and temporal reliability, and predictive and discriminant validity of the scale (Pratto, et al., 1994).
- The Self-Monitoring Scale (Snyder, 1974), a 25-item, true-false scale that assesses individual differences in expressive self-presentation and impression management. There is ample evidence for the internal and temporal reliability of the scale, along with its predictive validity (Lennox & Wolfe, 1984; Snyder, 1974).

The GEQ is a commonly used scale to measure acculturation and ethnic identity, and was included as a manipulation check for ethnic/cultural differences. It contains 38 statements, 25 rated on a 5-point Likert scale from strongly disagree to strongly agree and 13 rated on a 5-point scale from very much to not at all. The target group mentioned in the GEQ was modified to be applicable to each ethnic group. Analyses of the GEQ Total score, which was the mean of all items after reverse coding those negatively loaded, indicated that the Chinese sample had significantly higher scores than American born Chinese reported by Tsai, et al. (2000), $t(42) = 3.04$, $p = .004$, $d = .46$, demonstrating that our Chinese sample identified themselves as Chinese and strongly with Chinese culture more so than American born Chinese. GEQ norms for Hispanics do not exist but their scores were comparable to the Chinese in our sample.

Stakes

Many studies in the deception literature have examined lies produced in situations in which participants were not very motivated one way or another to lie or tell the truth because they did not believe there were rewards or punishments associated with their performances. Higher-stakes studies are more analogous to real-life situations that face law enforcement and security personnel, and the behavioral indicators associated with veracity and lying that emerge from higher-stakes studies are different and more compelling than those from lower-stakes studies (DePaulo, et al., 2003; Frank & Svetieva, 2013). Identifying indicators that are based in low-stakes studies that are not analogous to real-life situations, or that are otherwise not validated, and then training law enforcement personnel on them would have dire consequences. At least one study has demonstrated detrimental effects of training to detect lies when non-validated indicators are used (Kassin & Fong, 1999). Consequently in this study participants were led to believe that their statements would be read by investigators

who would make a determination about the believability of the statements, and that there were rewards and punishments for the participants depending on those determinations.

The stakes associated with their performances were as follows:

- If they lie about what they witnessed and wrote and are believed, they will receive an additional \$75 and will be allowed to leave early.
- If they lie about what they witnessed and wrote but are not believed, they will receive no additional money and will have to stay an additional hour filling out a long questionnaire.
- If they tell the truth about what they witnessed and wrote and are believed, they will receive an additional \$10 and will be allowed to leave early.
- If they tell the truth about what they witnessed and wrote but are not believed, they will receive no additional money and will have to stay an additional hour filling out a long questionnaire.

The stakes were different for lying and telling the truth because they reflected the stakes that occur in real life for the type of investigative interview examined in this study. Being a successful liar is likely associated with relatively large rewards in real life; and participants who may not be inclined to do so in the first place require additional motivation to do so. As it is easier for people to tell the truth, there are indeed rewards for telling the truth, but they are lower than when successfully lying. If, however, the type of investigative interview were different, different stakes might be more appropriate. For example, if participants were falsely accused of lying about what they had witnessed, there would be a larger stake in being perceived as truthful. This, however, would require a different study. Thus readers are cautioned to interpret the findings reported below vis-à-vis the particular way in which the experiment was conducted, including the stakes involved.

Procedures

Upon arrival to the laboratory, participants were led into a private instruction and consent area with a computer workstation. All forms, protocols, questionnaires, surveys, and instructions were presented in the native language of the participant (i.e., English, Spanish, or Chinese). After informed consent was obtained, participants were left alone to complete the pre-session measures via web-based surveys on the computer. When completed, participants rang a bell to call the experimenter back into the room. The instructions were then presented to the participants via audio PowerPoint.

The instructions informed participants that they will view a video that may or may not portray a crime, and then be asked to write two statements about what they witnessed.

Participants were not specifically told at this time that they would be writing both a false a true statement and in what order. Separate pilot testing of the procedures indicated an order effect when participants are asked to produce true and false statements. Participants in that pilot study who wrote a true statement first and then a false one reported that the content of their false statements was heavily influenced by their knowledge of what they had just written in the true statement. Those who wrote a false statement first, however, reported that their false statements were uncontaminated by their knowledge of what they may have written in a true statement (because they wrote the false statement first). True statements written after a false statement were not affected by order. Moreover, uncontaminated false statements written first are more likely to have ecological validity. Thus we opted to have all participants write both statements in a fixed false-true order, which was at first unbeknownst to the participants.

Participants were told that one of their statements will be selected to be read and evaluated by experts, that they may also be interviewed about what they witnessed and wrote, and that a determination will be made about whether they are lying or telling the truth. They were informed about the stakes associated with their performance, as described above. They were told that there are big rewards if they are believed, but also serious consequences if they are not believed; thus they must try to be as convincing as possible. After the instructions were delivered, including the stakes, and any questions answered, participants rated the severity of the consequences they faced in the experiment using a 10 point scale. Mean ratings across all participants was 5.13, commensurate with the participants' perception of moderate level of stakes in the experiment.

After ensuring that participants understood the instructions and stakes, the first experimenter left and a second experimenter entered and escorted participants to a second room in which there was a computer workstation. The experimenter asked the participants to follow the on-screen instructions, which read as follows:

“You will now see a video that may or may not show a crime. There is no audio in the video. You should play the video only once and you cannot take notes. You can click the bottom right corner of the image to view the video in full screen. Click Next when you're ready to watch the video.”

Video 4 then played in its entirety, lasting 49 s. When the video finished, participants were then instructed to write a statement about what they saw in the video, as follows:

“Soon you will be interviewed by a security officer about what you witnessed. Before that interview, you first need to write a FALSE statement about what you

witnessed in the video. Please write this FALSE statement about what you have just witnessed knowing that the statement may be read by a security officer and others who will determine whether it is believable or not. You may also be interviewed based on your statement. Thus write this FALSE statement to be as believable as possible. Use the paper and pen provided. You can write as much or as little as you want.

Write this statement in your native language. Do NOT go back and review the video.

When you are done, please ring the bell and wait for the experimenter. Do NOT click Next.”

Lined paper and a pen were provided. The participant wrote the statement, and then rang a bell when done to call the experimenter back into the room. The experimenter ascertained that the participants understood the instructions (i.e., wrote a FALSE statement), and then took the statement and labeled it “A” in plain sight of the participants. The experimenter then instructed the participants to click to the next screen after the experimenter left the room. The following instructions appeared:

“Now please write a TRUE statement about what you witnessed in the video. Remember, this TRUE statement may be read by a security officer and others who will determine whether it is believable or not. You may also be interviewed based on your statement. Thus, write this TRUE statement to be as believable as possible. Use the paper and pen provided. You can write as much or as little as you want.

Write this statement in your native language. Do NOT go back and review the video.

When you are done please ring the bell and wait for the experimenter. Do NOT click Next.”

When participants rang the bell, the experimenter re-entered the room, ascertained that participants understood the instructions (i.e., wrote a statement a TRUE statement), and then took the statement and labeled it “B” in plain sight of the participants.

Immediately after the writing exercise was completed, one of the statements was indeed selected and participants were interviewed about their statements using a standard interview protocol (i.e., questions were designed a priori and administered uniformly to all participants). As the analysis of the interview is not part of this study no further mention of it will be made.

After completion of the interview the second experimenter escorted the participants back to the initial instructions and consent area with the first computer workstation. The experimenter exited and the first experimenter re-entered. Participants were asked to complete some brief post-session

measures on screen, were debriefed, compensated, and excused.

Coding

Structural Balance The number of lines devoted to the prologue, incident, and epilogue portions of each statement was counted, and statements were coded dichotomously as either balanced or unbalanced. Balanced statements were defined to contain at least 33% of the total lines in the statement, with at least 20% devoted to each of the prologue and epilogue.

Word Count The total word count for each statement was tallied.

Unique Sensory Detail (USD) and Spatial Detail (SD) The number of sentences in each statement that contained evidence for either/or both USD and SD – that is, specific descriptions generated by the five sensory perceptions to include sight, sound, touch, smell, taste and touch, or specific locations and the physical relationships of objects, people, etc., in relation to one another – was counted.

Emotion The number of sentences within the epilogue of the statement that contained words that described the writer's emotion was counted.

Extraneous Information Each sentence within a participant's response that contained extraneous information was identified, regardless of the extent of the extraneous information within that one sentence, and the total number of sentences within each statement was tallied.

Equivocation The number of words or phrases within each statement that were construed as equivocation words/phrases from the writer's vantage point was counted. Equivocation pertaining to actions by individuals in the video were not counted. For example, the statement "*The motorcyclist was sort of responsible for what happened*" was not counted as equivocation because the equivocation pertained to an individual in the video and not to the writer's perception of what happened. But the statement "*I sort of recall the vehicle being white*" was counted as equivocation because it pertained to the writer's perception.

Non-Prompted Negation (NPN) The number of words or phrases within each statement that were construed as NPN as they pertained to the writer was counted. For example, the sentence "*The motorcycle did not cross the road*" was not counted as NPN because it pertained to an action by an individual in the video and not the writer. The statement "*I did not see a motorcycle*" was

counted as NPN because it pertained to the writer's actions or perceptions.

Moderating Adverbs Each word that constituted an Editing, Minimizing, or Intensifying adverb within a response was identified, and the total number of instances within each statement was tallied for each of these three types of adverbs. Adverbs that were counted had to pertain to the actions or perceptions of the writer; adverbs that pertained to activity by the individuals in the video were not counted.

Passive Voice The number of uses of the passive voice within each statement was counted.

Coding Procedures and Reliability Two coders coded the linguistic features of the statements. One coder (VAS) had several decades of law enforcement experience and extensive experience in conducting statement analysis in real-life investigative settings, was fluent in English and Spanish and coded the English and Spanish statements. A second coder, also an individual with several decades of experience in a law enforcement agency, was fluent in English and Chinese and coded the English and Chinese statements. Both coders first independently coded statements from 20 randomly selected English statements (10 true and 10 false). Initial reliabilities (Kappa for Structural Balance, ICCs for all other categories) were calculated on the initial set of 20 statements and ranged from .74 to 1.00. The coders were then instructed to arbitrate any disagreements and recalibrate their codes. They then independently coded the transcripts and statements from a new set of 20 English statements. Reliabilities computed across all 40 statements coded were high and acceptable for all coding categories ($.79 < \text{interrater reliability} < 1.00$). Both coders then completed coding the remaining English statements, and then coded either the Spanish or Chinese statements. Statements were provided with no marks or indicators of condition.

When the writer made a very obvious typographical error and it was readily apparent from the context what the writer intended (e.g., "cor" instead of "car"), the misspelled word was analyzed and included in the word count (WC). If a determination about what the writer meant in the use of the misspelled word could not be made from the context, the word was still treated as a word for word count purposes but was not marked as any other applicable linguistic feature. Also, when writers crossed out words, phrases, or sentences and they could clearly be deciphered, they were analyzed for linguistic features and were included in the word count. If a determination about what the writer meant in the use of the crossed out words, phrases, or sentences

could not be made, they were not marked as any applicable linguistic feature nor were they included in the word count.

Results

Main Analyses

We first computed aggregate scores for the veracity and deception indicators by summing the codes for Emotions and USD-SD (veracity indicators), and the codes for Extraneous Information, Equivocation, NPN, Moderating Adverbs, and Passive Voice (deception indicators), separately for each statement. We then computed a Language (3) by Veracity Condition (2) by Indicator Type (2) mixed ANOVA on the aggregate scores. The Veracity Condition by Indicator Type interaction was significant, $F(1, 120) = 14.50, p = .000, \eta_p^2 = .11$. As predicted true statements had more veracity indicators than did false statements, while false statements had more deception indicators than did true statements (Figure 1). Importantly, the Language by Veracity Condition by Indicator Type was *not* significant, $F(2, 120) = 2.30, p = .10, \eta_p^2 = .04$, indicating that language did not moderate the interaction between Veracity Condition and Indicator Type.

In order to analyze each of the coded linguistic features separately, we computed a mixed-factor MANOVA using Language and Veracity Condition as independent variables and the scalar coded linguistic features as multiple dependent variables. The main effects of Language and Veracity Condition were significant, $F(14, 226) = 7.68, p = .000, \lambda = .45$; and $F(7, 112) = 3.90, p = .001, \lambda = .80$, respectively; the interaction was not, $F(14, 224) = 1.43, p = .15, \lambda = .84$. Follow up, univariate analyses decomposing the Veracity Condition effect indicated that true statements contained more elements of USD-SD than did false statements, while false

statements contained more minimizing adverbs than did true statements (both $ps < .05$; Table 2). For Structural Balance, we computed one-sample binomial tests on the proportion of statements that were coded balanced or unbalanced, separately for the true and false statements. As predicted a greater proportion of the false statements (77%) were coded as unbalanced, $z = 5.95, p = .000$; for true statements there were no differences between the proportion of statements coded as balanced (55%) or unbalanced, $z = 1.09, p = .28$.

Post Hoc Analyses

To decompose the significant main effect of Language, we computed follow-up, oneway analyses with pairwise comparisons using Bonferroni corrections. Chinese writers used more words, sensory and spatial detail, and intensifying adverbs than did the English and/or Spanish samples, while the English writers used more words conveying extraneous information (Table 3).

General Discussion

As a whole the selected linguistic features discriminated between true and false witness statements and the effect sizes were relatively large. Univariate analyses indicated that structural balance, unique sensory and spatial detail, and minimizing adverbs were particularly important in differentiating true and false statements. Importantly, language did *not* moderate the relationship between veracity and the coded features.

The findings were not produced without limitations, the first of which concerned the laboratory setting in which the crimes were viewed. Actually witnessing a hit and run occur,

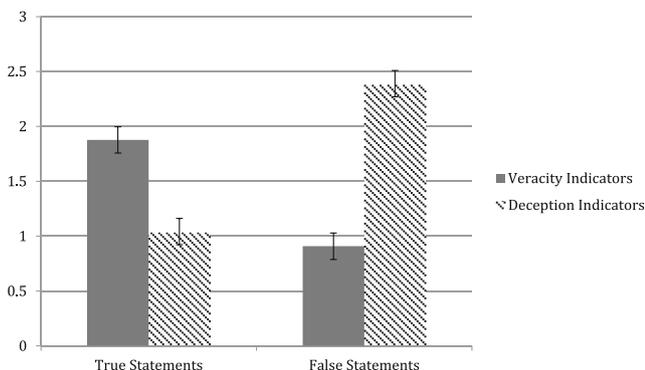


Fig. 1 Differences between True and False Statements as Measured by Veracity and Deception Indicators (error bars refer to Standard Errors)

Table 2 Means Standard Deviations (in parentheses), and Cohen's d for each of the Coded Linguistic Features, Separately for True and False Statements

Coded Feature	True Statements	False Statements	Cohen's d
Word Count	97.05 (62.59)	92.12 (74.05)	-.10
USD-SD	2.32 (1.61)	1.81 (1.51)	.46
Emotion	.11 (.31)	.10 (.44)	.02
Extraneous Information	.08 (.31)	.07 (.25)	.06
Equivocation	.14 (.54)	.17 (.55)	-.06
Non-Prompted Negation	.06 (.30)	.07 (.43)	-.03
Editing Adverbs	.02 (.13)	.02 (.16)	-.05
Minimizing Adverbs	.00 (.00)	.04 (.24)	-.17
Intensifying Adverbs	.62 (1.29)	.69 (1.45)	-.06
Passive Voice	.02 (.13)	.00 (.00)	.13

Note: USD-SD – Unique Sensory Detail and Spatial Detail

Table 3 Means, Standard Deviations (in parentheses), and Results of Pairwise Tests of Language Differences with Bonferroni Corrections

Variable	Chinese	English	Spanish	Bonferroni Results
Word Count	233.30 (154.13)	162.45 (110.76)	160.40 (101.32)	Chinese > English
USD-SD	5.49 (2.81)	4.05 (2.92)	2.62 (2.36)	Chinese > Spanish
Emotions	.33 (.78)	.24 (.49)	.05 (.31)	
Extraneous Info	.00 (.00)	.45 (.79)	.02 (.15)	English > Chinese, Spanish
Equivocation	.42 (.98)	.45 (1.43)	.05 (.22)	
Non-Prompted Negation	.14 (.64)	.21 (.58)	.05 (.22)	
Editing Adverbs	.09 (.37)	.03 (.16)	.00 (.00)	
Minimizing Adverbs	.07 (.34)	.05 (.23)	.00 (.00)	
Intensifying Adverbs	3.12 (3.41)	.63 (1.20)	.00 (.00)	Chinese > English, Spanish
Passive Voice	.00 (.00)	.05 (.23)	.00 (.00)	

Note: USD-SD – Unique Sensory Detail and Spatial Detail

with the associated sensory perceptions (e.g., sights, sounds, smells, etc.) would have been more realistic. Also, the task asked of the participants (i.e., to write a false statement) may have been too general. If individuals actually had to lie about what they witnessed, they would do so for a specific reason, such as wanting to conceal the identity of the driver of the car that did the hit and run, or needing to deny being in that particular place at that particular time. We purposely did not give these kinds of instructions to the participants because if we did, truth tellers would also have to lie because they would have had to put themselves into the hypothetical situation that we instructed them. This would have resulted in the true statements not really being true. Thus we gave participants the flexibility about exactly how to craft their false statement. The limitation in doing so, however, is that some participants only changed a seemingly minor or irrelevant detail in the false statement, which was likely not the type of statements one would obtain if one had a very specific reason for lying about having witnessed a crime, and which may not require the kinds of linguistic or grammatical choices when producing more realistic false statements. These limitations probably led to a dilution of the quality of the statements we obtained and analyzed. On one hand, this may have influenced the degree to which many of the linguistic features were observable or not in the statements, which in turn resulted in many near-zero categories of data (which was indeed observed). On the other hand, the limitations may have also made it more difficult to obtain statistical significance, which would be an acceptable Type II error. The fact that we obtained positive findings even with such statements obtained in such contexts may have been even more meaningful.

Another limitation concerned the specific features selected for coding in this study. As mentioned previously, we selected only a few SA categories that were deemed appropriate in this study given the experimental context and procedures, and for which there was previous empirical support. Different coded features may lead to the different outcomes.

A final limitation may have occurred because of the fixed order of the tasks in the study. It may have been possible, for example, for participants to provide inaccurate information in their true statements because of a bias of having written a false statement prior to writing the true statement. To mitigate against this possibility, we reviewed all the statements provided by the participants. We confirmed that all true statements were indeed accurate depictions of what occurred in the crime video they witnessed (albeit with large individual differences in the amount of details reported). We also confirmed that all false statements did indeed contain some kind of false information. Thus this bias for the potential recall of inaccurate information in the true statements did not occur in our study.

Regardless of these limitations the findings provided exciting initial evidence for the potential cross-cultural, cross-language generalizability of SA to differentiate truths from lies. That the SA features did indeed differentiate true and false statements was not new to the literature; the unique contribution of this study was the fact that multiple languages were tested in the same study and that language did not moderate the ability of the SA features to differentiate truths from lies. Although a number of studies have examined the ability of linguistic features to do so in different languages separately (Masip, et al., 2012; Ruby & Brigham, 1997; Schelleman-Offermans & Merckelbach, 2010), this study is the first to examine multiple languages in the same experiment, providing for an apples-to-apples comparison across languages and findings.

As mentioned earlier, cross-language consistency in the relationship between SA features and veracity suggests a potential pancultural similarity in the underlying psychological effects of lying, and similarity in the linguistic choices that mark those effects. Although it is not known whether memory is structured similarly across cultures, the current findings suggest the existence of a possible universal mechanism underlying the psychological demands placed on individuals when lying, and in the linguistic and grammatical choices that

individuals make when lying. If the rules of grammar and deep structure of language (Chomsky, 1957, 1972) and the principles of memory and recall (Undeutsch, 1989) are similar across cultures, and if there is cross-cultural similarity in the psychological demands placed on individuals when lying, then verbal indicators of truths and lies may occur regardless of culture, ethnicity, and language.

Cross-language similarity in the applicability of SA features has important potential practical implications. For example, the U.S., as many countries of the world, is a multi-ethnic, multi-language society, and a popular destination of peoples from many other countries of the world. Knowing that the principles and techniques of SA are applicable across cultures and languages would be a major boon to those individuals whose jobs rely on making determinations about truths and lies – such as law enforcement officers, customs and immigration officials, or airport security personnel – who interact with people speaking different languages on a daily basis.

It is interesting to speculate about why some of the coded SA categories did not differentiate true and false statements. It is very likely that some of the limitations discussed above concerning the nature of the task and the setting within which the crime was viewed limited the nature of the statements produced. Making the task more realistic and personal for the writers in the future may help to address this issue and allow for a fairer test of the various SA categories to differentiate truths from lies.

Although language did not moderate the relationship between SA features and veracity, there were some interesting language main effects. These effects have also been reported in a study of SA features coded in interviews and written statements in an experiment using a mock crime scenario (Matsumoto, et al., 2013). Although these findings clearly need to be replicated, they suggested cultural differences in the use of language that facilitate the use of some grammatical features but not others. These differences likely contribute to differences in overall communication styles.

Future research will need to replicate the findings obtained in this study in different contexts, with different languages, and different sources of statements (e.g., oral vs. written). In particular examining cross-linguistic differences in truths and lies related to one's own actions, not as a witness to someone else's actions, in a context in which participants are motivated and there are stakes for performance, may be a key test of the replicability of the findings reported here. Examining individual differences in the use of language across cultures in deceptive situations is also warranted.

Acknowledgments The authors would like to thank the following individuals for their aid in the collection of data in Study 1: Jeff Fujimoto for his aid in collecting data in Ecuador, Angel Avendano for his aid in collecting data in Bolivia, and Xiaohang Feng for her aid in collecting data in China. This report was prepared with the support of research grant FA9550-11-1-0306 from the U.S. Air Force Office of Scientific Research.

References

- Adams SH (1996) Statement analysis: What do suspects' words really reveal? *FBI Law Enforcement Bulletin* 65:12–20
- Adams SH, Jarvis JP (2006) Indicators of veracity and deception: An analysis of written statements made to police. *Speech, Language, and the Law* 13:1–22
- Chomsky N (1957) Syntactic structures. Walter de Gruyter, New York
- Chomsky N (1972) *Language and mind*. Harcourt, Brace, and Johanovich, New York
- Christie R (1970) Scale construction. In: Christie R, Geis FL (eds) *Studies in Machiavellianism*. Academic Press, New York, pp 10–34
- Connelly S, Allen MT, Ruark GA, Kligyte V, Waples EP, Leritz LE, Mumford MD (2006) Exploring content coding procedures for assessing truth and deception in verbal statements, Year 3, Cumulative Final Report. University of Oklahoma, Norman, OK
- Costa PT, McCrae RR (1992) Revised Neo-Personality Inventory (NEO-PI-R) and Neo Five Factor Inventory (NEO-FFI). Psychological Assessment Resources, Odessa, FL
- DePaulo BM, Lindsay JJ, Malone BE, Muhlenbruck L, Charlton K, Cooper H (2003) Cues to deception. *Psychological Bulletin* 129: 74–118
- Frank MG, Svetieva E (2013) Deception. In: Matsumoto D, Frank MG, Hwang HS (eds) *Nonverbal Communication: Science and Applications*. Sage, Newbury Park, CA, pp 121–144
- Hauch, V., Blandon-Gitlin, I., Masip, J., & Sporer, S. L. (2012, 23–27 April 2012). *Linguistic cues to deception assessed by computer programs: A meta-analysis*. Paper presented at the Proceedings of the EAACL 2012 Workshop on Computational Approaches to Deception Detection, Avignon, France.
- Johnson MK (1988) Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General* 117: 390–394
- Johnson MK, Raye CL (1981) Reality monitoring. *Psychological Review* 88:67–85
- Johnson MK, Raye CL (1998) False memories and confabulation. *Trends in Cognitive Sciences* 2:137–145
- Kassin SM, Fong CT (1999) "I'm innocent!": Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior* 23:499–516
- Lennox RD, Wolfe RN (1984) Revision of the Self-Monitoring Scale. *Journal of Personality and Social Psychology* 46:1349–1364
- Masip J, Bethencourt M, Lucas G, Sanchez-San Segundo M, Herrero C (2012) Deception detection from written accounts. *Scandinavian Journal of Psychology* 53:103–111
- Masip J, Sporer SL, Garrido E, Herrero C (2005) The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime, and Law* 11: 99–122
- Matsumoto D, Hwang HC, Sandoval VA (2013) Ethnic similarities and differences in linguistic indicators of veracity and lying in a moderately high stakes scenario. *Journal of Police and Criminal Psychology*. doi:10.1007/s11896-013-9137-7
- McCrae RR, Costa PT (1997) Personality trait structure as a human universal. *American Psychologist* 52:509–516
- McCrae RR, Terracciano A, Khoury B, Nansubuga F, Knezevic G, Djuric Jovic D (2005) Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology* 88:547–561
- Newman ML, Pennebaker JW, Berry DS, Richards JM (2003) Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin* 29:665–675
- Paulhus DL, Williams KM (2002) The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality* 36:556–563

- Porter S, Birt AR, Yuille JC, Lehman DR (2000) The negotiation of false memories: The influence of interviewer and rememberer characteristics on memory distortion. *Psychological Science* 11:513–516
- Porter S, Yuille JC (1996) The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior* 20:443–459
- Pratto F, Sidanius J, Stallworth LM, Malle BF (1994) Social Dominance Orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology* 67:741–763
- Rabon D (1994) *Investigative Discourse Analysis*. Carolina Academic Press, Durham, NC
- Ruby CL, Brigham JC (1997) The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations: A critical review. *Psychology, Public Policy, and Law* 3:705
- Rudacille WC (1994) *Identifying lies in disguise*. Kendall-Hunt, Dubuque, IA
- Sapir A (1996) *The laboratory for scientific interrogation course on scientific content analysis (SCAN) workbook*. Laboratory for Scientific Interrogation, Phoenix, AZ
- Schafer, J. R. (2007). *Grammatical differences between truthful and deceptive written narratives*. Unpublished doctoral dissertation.
- Schelleman-Offermans K, Merckelbach H (2010) Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling* 7:247–260
- Snyder M (1974) Self-Monitoring of Expressive Behavior. *Journal of Personality and Social Psychology* 30:526–537
- Sporer SL (2004) Reality monitoring and detection of deception. In: Granhag PA, Stromwall LA (eds) *The detection of deception in forensic contexts*. Cambridge University Press, New York, pp 64–80
- Tsai JL, Ying Y-W, Lee PA (2000) The meaning of "being Chinese" and "being American": Variation among Chinese-American young adults. *Journal of Cross-Cultural Psychology* 31:302–332
- Undeutsch U (1989) The development of statement reality analysis. In: Yuille JC (ed) *Credibility Assessment*. Kluwer Academic/Plenum Publishers, New York, pp 101–119
- Vrij A (2007) Criteria based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law* 11:3–41
- Vrij A, Mann S (2006) Criteria-based content analysis: An empirical test of its underlying process. *Psychology, Crime, and Law* 12:337–349
- Weiner M, Mehrabian A (1968) *Language within language: Immediacy, a channel in nonverbal communication*. Appleton-Century-Crofts, New York
- Weintraub W (1989) *Verbal behavior in everyday life*. Springer, New York
- Zaparniuk J, Yuille JC, Taylor S (1995) Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry* 18:343–352