Journal of Investigative Psychology and Offender Profiling
J. Investig. Psych. Offender Profil. **13**: 56–69 (2016)
Published online 22 July 2015 in Wiley Online Library (wileyonlinelibrary.com). **DOI**: 10.1002/jip.1442

Linguistic Cues of Deception Across Multiple Language Groups in a Mock Crime Context

HYISUNG C. HWANG[1,2,*], DAVID MATSUMOTO[1,2] and VINCENT SANDOVAL[3]

[1]*San Francisco State University, San Francisco, CA, USA*
[2]*Humintell, El Cerrito, CA, USA*
[3]*Federal Bureau of Investigation (retired), Washington D.C., WA, USA*

Abstract

A recent study showed that specific linguistic and grammatical features of a technique commonly referred to as statement analysis are applicable across different language groups. One limitation of that study was that it used an eyewitness crime video paradigm, which might be different from writing a statement after committing an actual criminal act. We remedied that limitation by using a mock crime paradigm. In this study, three language groups (English, Spanish, and Chinese) produced statements after committing a mock crime, taking a check, in an experimental context. Certain linguistic features significantly discriminated truths from lies similarly across the different language groups, suggesting that statement analysis might be applicable as a reliable indicator of deception across languages. Copyright © 2015 John Wiley & Sons, Ltd.

Key words: language; statement analysis; ethnicity; deception detection; crime

Understanding linguistic and grammatical cues from written statements is one way to distinguish truths from lies. This method is critically important in investigative contexts because investigators commonly require written statements from suspects and witnesses. To increase the efficacy of the interview, the ability to analyse the credibility of statements provided is useful for interrogators and interviewers who must discern truths from lies. However, determining the veracity of written statements is complicated, because writing allows people to have time to develop convincing stories, which may reduce the risks of detection. Fewer written linguistic or grammatical indicators of deception are likely to be present when liars write well-prepared narratives. Thus, utilising proper techniques with a deep understanding is crucial.

One class of techniques for analysing written statements for veracity and deception is known as statement analysis (SA; also known as scientific content analysis, investigative

*Correspondence to: Hyisung C. Hwang, Department of Psychology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA.
E-mail: hwang@sfsu.edu

discourse analysis; Leo, 2008). SA is an effective technique that can be used to guide investigative interviews by analysing the words people use (Vrij, 2008; 'Statement Analysis', n.d.). It is a broad concept that includes specific systems such as criteria-based content analysis (CBCA; Ruby & Brigham, 1997) and reality monitoring (RM; Johnson & Raye, 1998). SA is rooted in psycholinguistic research from the early 1900s; its more modern forms stem from the work of Undeutsch (1989) and a technique known as statement validity analysis, which was based on the premise that statements associated with actual memories differ from those based on fabrication or fantasy (Undeutsch, 1989).

Many studies have been conducted to examine the validity of different SA techniques to detect deception (e.g. Duran, Hall, Mccarthy, & Mcnamara, 2010; Masip, Bethencourt, Lucas, Sanchez-San Segundo & Herrero, 2012; Porter & Yuille, 1996; Ruby & Brigham, 1997; Sporer & Schwandt, 2006). CBCA is one of the most studied strategies; it has 19 criteria such as general, unusual, motivational, and stylistic features (Undeutsch, 1954), which can be flexible depending on usage. Higher numbers of the criteria in a statement indicate a higher probability that the statement is truthful (Colwell, Hiscock-Anisman, & Fede, 2013). Willén and Strömwall (2012) found that some individual CBCA criteria indeed differentiated truths from lies.

Reality monitoring (Johnson, 1988; Johnson & Raye, 1998) has also received considerable empirical attention. RM refers to the cognitive operations associated with attributing memories to internal (fabricated) versus external (perceived) events and is based on the rationale that memories of true events differ in quality and content from fabricated memories in a number of ways (Johnson & Raye, 1981). Johnson and Raye (1981) believed that more external-sensorial information and contextual information would appear in memories of actually experienced events, because these are encoded in memory when events actually occur in reality. Conversely, such sensory and contextual information should occur less frequently in false accounts of memories. RM-based techniques have led to accuracy rates in the 80% range when predicting statements as honest or deceptive (Masip *et al*., 2005).

Despite the evidence for its validity and potential operational utility, however, SA has been criticised because of the lack of adequate evidence as to its application in various languages, as most empirical evidence has been derived from the original languages (e.g. German and English) in which it was developed (Leo, 2008). Considering the potential efficacy of SA, there is a great possibility and need to test and ascertain its utility across languages. Extending the usability of SA to various languages is demanding but meaningful as it is one way to examine the reliability of SA.

To be sure, a few studies have examined SA indicators of deception in languages other than English (Masip *et al*., 2012; Ruby & Brigham, 1997; Schelleman-Offermans & Merckelbach, 2010; Spence *et al*., 2012). One limitation of these studies, however, was that each studied a single different language and no one study compared different languages within the same study using the same methodology. Thus, although these earlier studies were suggestive of the potential cross-language applicability of SA, comparing results across them is difficult because study differences confound the languages examined.

A more recent study addressed this limitation (Matsumoto, Hwang, & Sandoval, 2014). In that study, 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 Spanish and Chinese language groups were selected as they are the largest foreign language groups amongst immigrants

DOI: 10.1002/jip

and in the overall population in the US. Selected SA linguistic features (e.g. unique sensory detail and spatial detail (USD–SD) and minimising adverbs, described below) discriminated between true and false witness statements at statistically significant rates. More importantly, language did not moderate the relationship between veracity and the coded features.

This latest study described above contributed to the scientific evidence by showing that specific and reliable linguistic and grammatical features of SA were applicable across multiple language groups. However, that study was also limited, because writing about having witnessed a crime video may differ from actually experiencing and committing a criminal act. Watching a crime video could be likened to watching a television drama or playing a video game, and the artificiality associated with such tasks may have reduced the quality of the statements produced and analysed. To extend that study and to remedy this particular limitation, we examined the cross-language applicability of SA in this study using a mock crime paradigm. One reason for using a mock crime scenario is that the literature has emphasised the importance of stakes and motivation in lying in experimental contexts (DePaulo *et al*., 2003; Frank & Svetieva, 2013; Matsumoto & Hwang, 2014). In the current study, participants from three language groups (English, Spanish, and Chinese) produced statements after committing a mock crime, taking a check, in an experimental context. The written statements provided by the participants were analysed using SA.

LINGUISTIC MARKERS OF VERACITY AND LYING USED IN THIS STUDY

In this study, several of the same SA categories that were tested in previous studies (Matsumoto, Hwang, & Sandoval, 2013, 2014) were used: USD–SD, extraneous information, equivocation, non-prompted negation, passive voice, and moderating adverbs (descriptions below come from both Matsumoto *et al*., 2013 and 2014).

Indicators of veracity

Unique sensory detail and spatial detail

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, and so on in relation to one another (Adams & Jarvis, 2006). Truthful statements are expected to contain USD–SD details about a specific event. CBCA (Porter & Yuille, 1996; Undeutsch, 1989; Vrij, 2007) and the RM frameworks (Johnson, 1988; Johnson & Raye, 1981) 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. Matsumoto *et al*. (2014) also found this to be true.

Indicators of lying

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

DOI: 10.1002/jip

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). This idea has been supported by many studies (DePaulo *et al*., 2003; Matsumoto *et al*., 2013; Vrij, 2007)

Equivocation

Equivocation refers to vague, ambiguous language that may be used purposely to deceive. Equivocation words (e.g. maybe and kind of) 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). Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more equivocation when writing statements in English.

Non-prompted negation (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 ones (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 non-prompted negation (NPN) both when writing statements in English and in oral interviews.

Passive voice

When describing actions, people generally assume responsibility for those actions by employing the active voice. 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).

Moderating adverbs

Moderating adverbs consists of *intensifying adverbs* (e.g. very, really, and honestly), which are typically used when a communicator is attempting to convince another person of something; *minimising adverbs* (e.g. only and just), which are used to minimise the role of the actor; and *editing adverbs* (e.g. after, next, and so), indicating a temporal lacunae (Rabon, 1994; Schafer, 2007). Adverbs are often used to edit information that might be crucial to an inquiry. Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more moderating adverbs both when writing statements and in oral interviews.

OVERVIEW OF THE STUDY

Based on the recent findings demonstrating cross-language applicability of certain categories of SA as an indicator of veracity and deception (Matsumoto *et al*., 2014), we hypothesised that the coded SA categories would differentiate truthful statements from lying ones across the three languages tested.

DOI: 10.1002/jip

METHODS

Participants

All participants were adults aged 18 years or older and came from one of the three ethnic/language groups: European Americans, Chinese immigrants, and Hispanic immigrants. The European Americans were all born and raised in the US and whose first language was English ($n = 35$ for men, $n = 28$ for women; $n = 38$ for lie, $n = 25$ for truth). The Hispanics were individuals who were born and raised in any country in Central or South America or whose parents were born in any of those countries, and whose first language was Spanish ($n = 24$ for men, $n = 25$ for women; $n = 23$ for lie, $n = 26$ for truth). The Chinese participants were individuals born and raised in the People's Republic of China, Hong Kong, or Taiwan or whose parents were born and raised in those countries and whose first language was Mandarin or Cantonese ($n = 16$ for men, $n = 40$ for women, $n = 1$ for unidentified; $n = 24$ for lie, $n = 33$ for truth). As a manipulation check on language fluency, participants were asked to self-evaluate their reading and writing levels (poor to excellent) in the target language. Only participants who highly rated their reading and writing skills in the primary language were selected for participation. Additionally, participants' self-ratings of their ethnic group identities were checked via the General Ethnicity Questionnaire (GEQ; Tsai, Ying, & Lee, 2000; more below). Statements from participants who withdrew consent at the end of the experiment or misunderstood their condition or experimental roles (e.g. forgot to enter the file room and take the check, did not write in their primary language, or were confused with their assigned condition) were excluded.

Measures

Portions of the descriptions of the measures and procedures have been reported previously (Matsumoto & Hwang, 2014; Matsumoto *et al*., 2013). At the beginning of the experiment, all participants completed a series of questionnaires including a brief demographic questionnaire, the GEQ, the NEO-Five Factor Inventory (Costa & McCrae, 1992), the Social Dominance Orientation Scale (Pratto, Sidanius, Stallworth, & Malle, 1994), an adapted version of the Schwartz Value Scale (Schwartz, 2006), the Satisfaction with Life Scale (Diener *et al*., 1985), and the Self-Monitoring Scale (Snyder, 1974). Participants also completed an emotion checklist at the beginning and the end of the experiment. This checklist included 12 emotion words (guilt, fear, anger, embarrassment, worry, contempt, excitement, disgust, amusement, nervousness, surprise, and interest) rated on 9-point scales labelled 0 = *None*, 4 = *Moderate Amount*, and 8 = *Extremely Strong*.

The GEQ is a commonly used scale to measure acculturation and ethnic identity and was included as a manipulation check for ethnic/cultural differences. This questionnaire 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 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 our Chinese sample had significantly higher scores than the American-born Chinese and Chinese who immigrated to the US before the age of 12 reported by Tsai *et al*. (2000), $t(64) = 14.58$, $p < .001$, $d = .85$; $t(64) = 7.87$, $p < .001$, $d = .46$, respectively. These analyses demonstrated that our Chinese sample identified themselves as Chinese and very strongly with Chinese

DOI: 10.1002/jip

culture, more so than American-born Chinese. For the European and Spanish groups, we modified the GEQ questions with target terms and language (e.g. How much do you speak English at home?). For the European American GEQ, the GEQ total mean scores were compared with those for American cultural domains reported by Tsai and her colleagues (2000) and with European Americans reported in Tsai, Knutson, and Fung (2006). There were no differences in either comparison, $t(61) = -1.394$, $p = .169$. $d = -.176$, and $t(262) = .69$, $p = .49$, $d = .102$. Norms for Hispanics using this same measure do not exist, but their scores were comparable with the Chinese and Americans in our sample. All participants reported their first language as a target language.

Procedure

Pre-session

Upon arrival to the laboratory, participants were instructed about the study and completed the consent form. Participants first completed the pre-session measures in order to avoid any possible carryover effects of their experience in the experiment. They were then given detailed instructions on the experiment, which differed depending on their truth or lie condition. The truth condition required participants not to take a check made out to cash for $200 and to tell the truth in the interviews and written statement. The lie condition required participants to take the check and lie in the interviews and written statement. The assignments were determined randomly prior to the participants' arrival to the laboratory. Participants were told that they would be interviewed regarding what they did in the file room, where the $200 check was located, and that they would have to persuade the interviewers about their honesty. Participants were told that they would earn a minimum of $30 for their participation, and bonuses of anywhere from $0 to $50 depending upon their assigned condition and the judgments of the interviewers. In reality, all participants received a standard fee of $40. After the introduction, participants rated the severity of the aforementioned expected consequences if they were judged to be lying in the experiment (the expected consequence when they shall fail in convincing an interviewer), using a scale from 1, *No consequence, even slightly pleasurable*, to 10, *Maximum consequence, even slightly painful*. The overall mean was 5.95, and there were no differences between the three language groups, $F(2,164) = .426$, $p = .654$. These findings indicated that the expected seriousness that the participants had about the experiment was at least on a moderate level.

Interviews and statement

After the pre-session, participants were guided to an interview room for an initial screening interview, the purpose of which was to ascertain participants' intent to commit a crime. Once the first interview was completed, participants waited nearby and then entered the file room, where the check was located. Depending on participants' veracity condition, they stole the check or left it where it was. After the file room, participants were escorted to the next interview. Prior to the interview, the interviewer asked participants to write a statement, in their native language, about what they did in the file room. A pen and lined paper were provided. The interviewers left the interview room during the writing. Once participants finished their writing, they rang a bell and the interviewer re-entered the room and briefly reviewed it before starting the second investigative interview, asking standardised questions in order to investigate participants' veracity. The

DOI: 10.1002/jip

interviews were conducted in English. As the purpose of this study was to examine whether language moderated the ability of SA categories to differentiate true from false written statements, the interviews were not analysed, and no further mention of them will be made.

Post-session

After completing the interview, participants were escorted to a debriefing room and completed the post-session measures (the emotion checklist described previously). The aim of the experiment was explained and they were given the standardised compensation fee, $40, and no punishment.

Coding

The SA categories described earlier were coded as follows:

Unique Sensory Detail and Spatial Detail

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.

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 were counted.

Non-prompted negation

The number of words or phrases within each statement that were construed as NPN as they pertained to the writer was counted.

Passive voice

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

Moderating adverbs

Each word that constituted an editing, minimising, 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.

DOI: 10.1002/jip

Coding procedures and reliability

Statements were coded by two trained raters who were blind to the conditions of the participants in the experiment. One coder had several decades of law enforcement experience and extensive experience in conducting SA 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); the calibration and reliability check was conducted in English, which was the common language of coders. Initial reliabilities (intra-class correlations—ICCs) were calculated on the initial set of 20 statements and ranged from .89 to 1.00. The coders were then instructed to arbitrate any disagreements and recalibrate their codes. They then independently coded the statements from a new set of 20 English statements. Reliabilities computed across all 40 statements coded were high and acceptable for all coding categories ($.87 < \text{ICC} < 1.00$). One coder then completed coding the remaining English statements and then the Spanish statements; the other coder coded the 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') or if the writer crossed out words and the words were legible, the word was analysed for linguistic features. 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 coded for any applicable linguistic feature.

RESULTS

Main analyses

We computed descriptives of all SA variables (Table 1) and aggregate scores for the veracity and deception indicators by summing the codes for extraneous information,

Table 1. Descriptives for all SA variables

Variable	Truth Mean (*SD*)	Lie Mean (*SD*)
USD–SD	.99	.24
	(1.24)	(.62)
Extraneous information	.36	.60
	(.71)	(.88)
Equivocation	.55	.96
	(.80)	(1.09)
Non-prompted negation	.24	.34
	(.59)	(.55)
Passive voice	.00	.04
	(.00)	(.25)
Moderating adverbs	1.76	1.36
	(1.72)	(1.28)

USD–SD, unique sensory detail and spatial detail

DOI: 10.1002/jip

equivocation, NPN, moderating adverbs, and passive voice separately for each statement (USD–SD was used as the single veracity indicator). We then computed a language (3) by veracity condition (2) by indicator type (2) mixed three-way analysis of variance on the aggregate scores. The veracity condition by indicator type interaction was significant, $F(1, 157) = 9.827$, $p = .000$, $\eta_p^2 = .59$. As predicted, true statements had relatively more veracity indicators than did false statements, whilst false statements had more deception indicators than did true statements (see Figure 1 which reports residualised means in order to present the pure interaction effect between veracity condition and indicator type; Rosnow & Rosenthal, 1989). Importantly, the language by veracity condition by indicator type interaction was *not* significant, $F(2, 157) = .845$, $p = .431$, $\eta_p^2 = .011$, indicating that language did not moderate the interaction between veracity condition and indicator type.

In order to examine how individual SA variables differed as a function of veracity condition and participant's language, we computed an overall language (3) by veracity condition (2) multivariate analysis of variance (MANOVA) using the 6 SA variables as dependents. The main effect of veracity condition was significant, $\lambda = .874$, $F(6, 184) = 4.422$, $p < .000$, $\eta_p^2 = .126$. The main effect of language was also significant, $\lambda = .724$, $F(12, 368) = 5.370$, $p < .000$, $\eta_p^2 = .149$. There was no interaction of language and veracity condition, $\lambda = .899$, $F(12, 368) = 1.670$, $p = .074$, $\eta_p^2 = .051$.

To follow up the significant veracity condition main effect, we collapsed across languages and computed a logistic regression using veracity condition as the dependent variable and SA variables as covariates, using backward conditional exclusion criteria, in order to clarify which SA variables differentiated truthful and false statements. The final equation included three SA variables, USD–SD, extraneous information, and equivocation, and accounted for 68.7% overall correct classification of cases (Table 2). USD–SD, extraneous information, and equivocation significantly differentiated true statements from deceptive statements.

Post hoc analyses: gender differences and language

Suckle-Nelson *et al.* (2010) reported that women who responded deceptively were more aware of the need to keep their statement short and careful than were men who responded deceptively. Although Suckle-Nelson *et al.* did not use SA, it was possible that, regardless

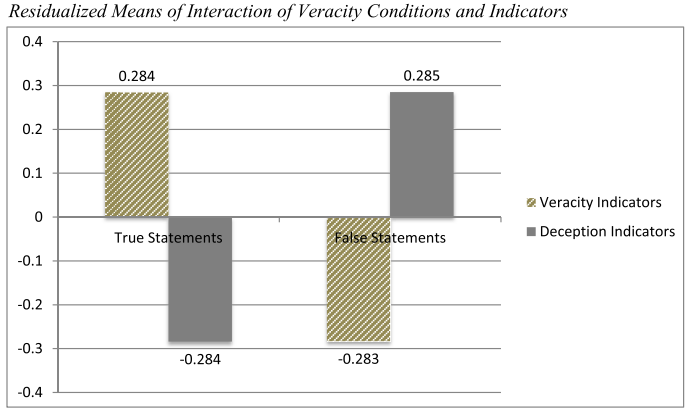

Figure 1. Residualised means of interaction of veracity conditions and indicators.

DOI: 10.1002/jip

Table 2. Final results of logistic regressions

Final model Chi-square	Overall correct classification (%)	False positive (%)	False negative (%)	Variables in	*B*	SE
$\chi 2(3, 163) = 34.172$, $p < .001$	68.70	14.72	16.56	USD–SD	1.035	.258
				Extraneous information	−.486	.243
				Equivocation	−.366	.195

USD–SD, unique sensory detail and spatial detail

of language, gender mattered. Thus, we conducted an overall MANOVA using language (3), veracity condition (2), and gender (2) as factors on the SA variables. There was no significant effect of gender, indicating that gender did not moderate the effects reported earlier.

We followed the main effect of language group reported earlier in the overall MANOVA by computing separate, univariate ANOVAs for each of the SA variables. USD–SD and moderating adverbs produced significant effects, $F(2, 189) = 15.779$, $p < .01$, $\eta_p^2 = .143$; $F(2, 189) = 21.776$, $p = .000$, $\eta_p^2 = .187$, respectively. We followed each of these with pairwise comparisons of the language groups using Bonferroni corrections. Chinese participants (mean [*M*] = .991, standard deviation [*SD*] = .124) produced more USD–SD than did the other two language groups ($M = .445$, $SD = .124$, and $M = .318$, $SD = .133$ for English and Spanish, respectively). A similar pattern was reported for moderating adverbs ($M = 2.256$, $SD = .189$; $M = 1.306$ $SD = .182$; and $M = .880$, $SD = .202$, for Chinese, English, and Spanish, respectively).

DISCUSSION

The findings supported the hypothesis that the SA features would distinguish truths from lies across languages. Specifically, the categories USD–SD, extraneous information, and equivocation were significant differentiators of veracity versus deception across multiple language groups. Participants telling the truth tended to write details, such as recalling particular scents, locations, and background noises or sounds, and to provide information directly relevant to the incident more than did liars when delivering truths in comparison to deceptive statements. This finding is consistent with the recent literature that tested the function of SA in eyewitnesses' statements about a crime (Matsumoto *et al*., 2014).

Some cautions, however, need to be exercised in interpreting these results. First, the study tested one type of crime (mock crime of theft) in a laboratory context. Thus, the application of our findings is limited to that crime type and context. Also readers have to be cautious in using the techniques of SA in actual cases, because it is possible that the results may vary with other types of crimes and in reality. Matsumoto and Hwang (in press) reported that people tended to perceive crimes such as a hit-and-run crime similarly across cultures, and the current findings vis-à-vis the function of SA were similar to the ones from the current study (Matsumoto *et al*., 2014). Thus, it is possible that the SA categories examined in both studies in the same three language groups might be reliably applicable at least for the cases of mock crime and witnessing a hit and run. Yet, the SA method may function differently with other types of crimes, and these should be studied in the future.

DOI: 10.1002/jip

Second, we tested participants who had no previous experience with actual investigative situations; thus, the findings were limited to people who are relatively naïve about investigative contexts (although there is an alternative possibility that people have indirect experiences through media or movies). The findings might be different with people who were already exposed to similar or real investigative contexts, and future studies may examine this possibility.

Third, the gender ratio of one of the three groups was not equivalent as the Chinese group had relatively fewer men. Statistically, the uneven sample sizes may have affected the reliability of our findings related to gender, despite the fact that our findings indicated no gender differences in the role of SA indicators across the three ethnic/language groups. This should be explored in the future.

Despite these limitations, this study made several important contributions. First, it extended the previous study by Matsumoto *et al*. (2014) to a different context. This is important because the previous study was limited in cases to witnesses who observed someone else's criminal act (hit and run) on video. Whilst in the previous study participants reported that they perceived the criminal act as serious, lying about someone else's crime may be different from lying about his or her own. This study extended the previous study from an eyewitness context to a criminal context in testing the role of SA. The reported findings are meaningful, because the data were derived from a context in which the participants had to actually commit a criminal act even though it was an experimental situation. Participants did, however, take stealing the $200 check seriously and became nervous about the act. Ideally, we need evidence based on data from actual criminals or witnesses in order to further test the usability of SA in detecting deception. The extension should provide interviewers and interrogators useful information to aid them in utilising SA with criminals.

Also, the current study replicated findings from many previous studies that have examined the various SA strategies with different categories (Porter & Yuille, 1996; Vrij, 2000). Specifically, extraneous information and equivocation were previously identified as important categories in SA (Matsumoto *et al*., 2014; Porter & Yuille, 1996; Vrij, 2000). This was supported by the current study across multiple language groups. Our finding added to the scientific evidence concerning the reliability of some features of SA in distinguishing truthful statements from lying ones across different crime contexts and languages.

The significant SA features tested in the study can possibly be applicable to actual investigative contexts. Our findings increased the possibility of SA as a pragmatic method in distinguishing truths from lies in statements and guided readers' attention to particularly useable SA categories as a constructive method in deception detection when analysing written statements provided by suspects or witnesses. As one of the customary or conventional processes of investigation, SA could be a valuable aid in making the investigation process effective.

Additionally, the study indicated that SA is applicable across at least three languages (English, Chinese, and Spanish) regardless of gender. This result is crucial, because at least in the US, crimes are committed by members of various ethnic and language groups. Dealing with non-English speakers and their statements in investigative contexts is not surprising or rare. Law enforcement officers or interrogators who may have to deal with written statements or use them as a source of interviews can possibly utilise the SA method with non-English speakers once the officers obtain the analytic skills. Considering that the major immigrant groups speak Spanish or Chinese, not only in the US, but around the world, the

DOI: 10.1002/jip

SA approach would be pragmatic to use particularly for those three language groups regardless of gender as well as for bilingual speakers amongst the three languages.

The language main effect was also interesting, showing that Chinese produced more USD–SD and moderating adverbs in general than did the other two language groups. This main effect finding was also reported in a previous study (Matsumoto *et al*., 2014). It is important to remember that this effect did not affect the veracity condition findings; that is, language differences in overall usage of these categories did not affect their ability to differentiate truthful statements from deceptive ones. The language effect, however, did suggest some real-world implications to the use of SA, as the greater use of these SA features in general may lead investigators to make incorrect inferences about veracity and deception.

The current study examined whether basic factors such as language and gender could affect the efficacy of SA in deception detection. However, there may be other elements that should be examined in order to distinguish truths from lies. For example, future studies will need to examine other languages (Arabic, French, etc.) using SA. Also, it would be interesting to test whether the current findings vary depending on different types of crimes and stakes levels. Ideally, collecting SA data from people, not only born and raised, but also currently living in that country and using the first language, would verify the pure usability of analysing linguistic information across different languages. Lastly, understanding how individual variation affects the reported findings would be valuable to explore.

ACKNOWLEDGMENTS

The authors wish to thank Ling-Ling Beerman for her assistance as a coder and Lauren Skinner, Melissa Vincent, Elizabeth Powers, and Sophia Nguyen for their assistance in collecting data used in this study.

DECLARATION OF CONFLICT OF INTEREST

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

FINANCIAL DISCLOSURE/FUNDING

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, S. H. (1996). Statement analysis: What do suspects' words really reveal? *FBI Law Enforcement Bulletin*, *65*, 12–20.

Adams, S. H., & Jarvis, J. P. (2006). Indicators of veracity and deception: An analysis of written statements made to police. *Speech, Language, and the Law*, *13*, 1–22.

DOI: 10.1002/jip

Colwell, J., Hiscock-Anisman, C., & Fede, J. (2013). Assessment criteria indicative of deception: An example of the new paradigm of differential recall enhancement. In B.S.

Connelly, S., Allen, M. T., Ruark, G. A., Kligyte, V., Waples, E. P., Leritz, L. E., & Mumford, M. D. (2006). *Exploring content coding procedures for assessing truth and deception in verbal statements, year 3, cumulative final report*. Norman, OK: University of Oklahoma.

Costa, P. T., & McCrae, R. R. (1992). Revised Neo-Personality Inventory (NEO-PI-R) and Neo Five Factor Inventory (NEO-FFI). Odessa, FL: Psychological Assessment Resources.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75.

Duran, N. D., Hall, C., McCarthy, P. M., & McNamara, D. S. (2010). The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics*, *31*, 439–462.

Frank, M. G. & Svetieva, E. (2013). Deception. In D. Matsumoto, M. G. Frank & H. S. Hwang (Eds.), *Nonverbal communication: Science and applications* (pp. 121–144). Newbury Park, CA: Sage.

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 EACL 2012 Workshop on Computational Approaches to Deception Detection, Avignon, France.

Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General*, *117*, 390–394.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.

Johnson, M. K., & Raye, C. L. (1998). False memories and confabulation. *Trends in Sciences*, *2*, 137–145.

Leo, R. A. (2008). *Police interrogation and American justice*. Cambridge, MA: Harvard University Press. ISBN 978-0-674-02648-3.

Matsumoto, D., & Hwang, H. C. (in press). Emotional reactions to crime across cultures. *International Journal of Psychology*. DOI: 10.1002/ijop.12103.

Matsumoto, D., & Hwang, H. C. (2014). Differences in word usage by truth tellers and liars in written statements and an investigative interview after a mock crime. *Journal of Investigative Psychology and Offender Profiling*. First published online 23 July 2014. DOI: 10.1002/jip.1423.

Matsumoto, D., Hwang, H. C., & Sandoval, V. A. (2013). Ethnic similarities and differences in verbal indicators of veracity and lying in a high stakes scenario. *Journal of Police and Criminal Psychology*. First published online 29 December 2013. DOI: 10.1007/s11896-013-9137-7.

Matsumoto, D., Hwang, H. C., & Sandoval, V. A. (2014). Cross-language applicability of linguistic features associated with veracity and deception. *Journal of Police and Criminal Psychology*. First published online 28 August 2014. DOI: 10.1007/s11896-014-9155-0.

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, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law*, *11*, 99–122.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*, 665–675.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*, 741–763.

Porter, S., Birt, A. R., Yuille, J. C., & Lehman, D. R. (2000). The negotiation of false memories: The influence of interviewer and rememberer characteristics on memory distortion. *Psychological Science*, *11*, 513–516.

Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, *20*, 443–459.

Rabon, D. (1994). *Investigative discourse analysis*. Durham, NC: Carolina Academic Press.

Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin*, *105*, 143–146.

Ruby, C. L., & Brigham, J. C. (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.

DOI: 10.1002/jip

Rudacille, W. C. (1994). *Identifying lies in disguise*. Dubuque, IA: Kendall-Hunt.
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.
Schwartz, S. H. (2006). Value orientations: Measurement, antecedents, and consequences across nations. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.), *Measuring attitudes cross-nationally – lessons from the European social survey* (pp. 161–193). Long: Sage.
Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*, 526–537.
Spence, K., Villar, G., & Arciuli, J. (2012). Markers of deception in Italian speech. *Frontiers in Psychology*, *3*, 453. DOI:10.3389/fpsyg.2012.00453.
Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, *20*, 421–446.
Statement Analysis. (n.d). In Wikipedia. Retrieved January 13, 2015, from http://en.wikipedia.org/wiki/Statement_analysis.
Suckle-Nelson, J. A., Colwell, K., Hiscock-Anisman, C., Florence, S., Youschak, K. E., & Duarte, A. (2010). Assessment criteria indicative of deception (ACID): Replication and gender differences. *Open Criminol. J.*, *3*, 23–30.
Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology*, *90*, 288–307.
Tsai, J. L., Ying, Y.-W., & Lee, P. A. (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. (1954). Die Entwicklung der gerichtspsychologischen Gutachtertatigkeit. In A. Wellek (Ed.), *Bericht uber den 19, Kongress der Deutschen Gesellschaft fur Psychologie* (pp. 1132–1154). Gottingen: Verlag fur Psychologie.
Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 101–119). New York: Kluwer Academic/Plenum Publishers.
Vrij, A. (2000). *Detecting lies and deceit*. Chichester, England: Wiley.
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. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed. pp.). Chichester, England: John Wiley & Sons.
Weintraub, W. (1989). *Verbal behavior in everyday life*. New York: Springer.
Willén, R. M., & Strömwall, L. A. (2012). Offenders' uncoerced false confessions: A new application of statement analysis? *Legal and Criminological Psychology*, *17*, 346–359. DOI:10.1111/j.2044-8333.2011.02018.x.

DOI: 10.1002/jip