Can reduced use of pronouns during deceptive versus truthful speech be observed in a language other than English?

Gina Villar¹, Joanne Arciuli¹, Alessio Barsaglini¹²

¹ Discipline of Speech Pathology, The University of Sydney, Lidcombe, Australia
² Department of Psychology, University Degli Studi di Padova, Padova, Italy

gvil2424@uni.sydney.edu.au, joanne.arciuli@sydney.edu.au, a.barsa@hotmail.it

Abstract

Previous research has identified linguistic markers of deception in English-speaking participants. Questions remain regarding the utility of such indicators in languages other than English. This study investigated the utility of pronouns as a linguistic marker of deception in Italian. Twenty native speakers of Italian gave their true and false opinions of social issue topics during an audio-taped interview. Analyses revealed that participants used fewer pronouns when they were lying compared to when they were telling the truth. These results suggest that linguistic features that have been associated with deception in English may also be useful indicators in other languages.

Index Terms: psycholinguistics, deception, pronouns, cross-linguistic

1. Introduction

Lying is a universal activity, with deceptive behaviour reported in every culture that has ever been examined [1]. Despite this, most of what we know about lying behaviour has been gleaned from the study of English-speaking participants. It is of value for scientific, commercial, political, forensic, medico-legal and military purposes to question whether some of the linguistic variables implicated in research on English serve as relevant indicators of deception in other languages. In the present study, we examined whether decreased pronoun usage is a reliable indicator of deception in the speech of native speakers of Italian.

Previous cross-cultural research on deception can be loosely grouped into five main areas. These include cross-cultural beliefs about reliable deception cues (e.g., many cultures incorrectly believe gaze aversion is a reliable indicator of deception [2]); differences between cultures in the evaluations of lying (e.g., see [3]); cross-cultural lie detection accuracy (e.g., see [4]); the effect of bilingualism on the experience of lying (e.g., see [5]); and lastly, cross-cultural differences in indicators of deception. Of the few studies in the latter group, most have focused on non-verbal behavioural cues (e.g., see [6]). At the time of writing there are no published journal articles (to our knowledge) that have considered linguistic indicators of deceptive speech in languages other than English.

It has been noted that the same reasons why liars give themselves away (such as heightened emotion, increased cognitive effort or attempts at behaviour control) may feasibly apply to most human beings, regardless of cultural or linguistic background [8]. Of the linguistic cues to deception that have been implicated in English-speaking participants, there are several that would appear to be amenable to application in research on languages other than English. Pronoun use appears to be one such linguistic marker [9].

A traditional linguistic definition of a pronoun as substituting for a noun or noun phrase might not accurately represent the structure and function of all pronouns [10]. For the purposes of this article we adopt a broad definition of a pronoun as a part-of-speech which designates a referent who (or which) is named or understood from the context. Pronouns can be classified into several types, with the most common being personal (e.g., I, me), possessive (e.g., hers, his), demonstrative (e.g., this, these), relative (e.g., who, whom), indefinite (e.g., all, any), reflexive (e.g., myself, themselves) and interrogative (e.g., which, what). Despite accounting for only .06% of the average English speaker’s word vocabulary, pronouns represent around 15% of the total number of words used in speech [11].

Linguistic markers of deception, such as pronoun use, offer the key advantage of being well-suited to automatic parsing of transcribed speech (as just like any other grammatical category, pronouns can be identified and counted using basic part-of-speech tagging systems). Thus, once transcribed, pronoun use can be systematically and objectively identified and tracked, for later comparison with baseline speech samples from the same individual.

Here we considered the utility of pronouns in discriminating between deceptive versus truthful Italian speech. While the lexical forms differ at a vocabulary level (e.g., the English word “I” versus the Italian word “io”) there are many commonalities across languages at a grammatical category level. For instance, all languages use pronouns, although there may be differences in the proportion of total words they account for in speech and text. Null languages such as Italian, Spanish, Portuguese and Romanian, permit the voluntary omission of pronouns in a sentence because they are generally considered redundant when their presence is implied by conjugation of the verb in the sentence. Consequently, subject pronouns are often omitted, however they may be included at the discretion of the speaker or writer, particularly for emphasis. Pronoun use may differ within a language: consider different forms of address pronouns used during formal versus informal communication contexts. Despite these divergences, pronoun use serves a similar communicative function across languages: to avoid repetition of nouns. Effective pronoun use minimises ambiguity, and thereby enhances understanding between speaker and listener.

Pronoun use represents a potentially productive source of investigation into changes in language behaviours during deception compared to truth-telling because it appears to vary according to the psychological state of the speaker [9]. For instance, individuals with depression use more personal pronouns than those without [12], while anxiety is negatively associated with personal pronoun use [13]. Moreover, pronouns appear to be affected by the mechanisms of speech....
production. For instance, a recent study found that fewer pronouns were produced following an initial disfluency (false start, repetition, speech error and so on), compared to fluent speech, perhaps in response to other cognitive demands involved in speech production [14]. It is widely accepted that lying can be an anxiety-evoking experience [15], while fabricating a lie and sustaining the deception requires considerable cognitive effort on the part of the liar [16]. Thus, it is feasible that both of these factors, either individually or in combination, may result in reduced pronoun use during deception compared to truth-telling.

Previous studies that have considered the utility of pronouns as markers of deception in English speech or written text have generally limited their examinations to one type: personal pronouns. Specifically, first person pronouns ‘I’, ‘me’, and ‘my’ (self-references) and third-person pronouns ‘he’, ‘she’, ‘him’, ‘her’, ‘they’, ‘them’ (other-references) have been the targets of investigation. Empirical data from several studies of English has been characterised by a reduction in the use of self-references along with a reduction in the use of third person pronouns during deception (e.g., see [17], [18]). The primary purpose of effective pronoun use in speech is to reduce ambiguity, and in turn to increase comprehensibility for the listener, a goal that is not exclusive to the use of first and third person pronoun. Thus, we suggest that measuring all types of pronouns, not just personal pronouns, is a worthwhile endeavour in deception research.

We expected that pronouns would appear less frequently in the deceptive versus truthful speech of speakers. There are two possible reasons for this. Firstly, lying is a cognitively effortful and anxiety-evoking process, conditions under which pronoun use has been shown to be reduced. Secondly, the primary purpose of pronoun use is to reduce ambiguity. We hypothesised that decreasing ambiguity would be aversive to liars, either consciously or unconsciously. Then again, it is an open empirical question as to whether the same kinds of effects seen in English would transfer to another language, such as Italian.

2. Method

2.1 Participants

The participants were 12 female and 8 male adult native speakers of Italian (N = 20). They were recruited in Sydney, Australia, through an advertisement in a local Italian newspaper, through flyers posted in Italian community-based organisations and through word of mouth. The participants did not receive any financial or other rewards for participation. The mean age was 56.8 years (SD = 14.57), the highest level of educational attainment was university, and participants had lived in Australia an average of 26.4 years (SD = 19.6).

2.2 Procedure

We employed a false opinion paradigm based on the procedure described by Frank and Ekman [19] and used in recent studies such as [20]. All recruitment notices described the study as an investigation of communication skills relating to social issues. To avoid inadvertently attracting a biased sample of either successful liars who are attracted to the challenge of deceiving an experimenter or unsuccessful liars who wish to improve their lying skills [21], the word ‘deception’ was not mentioned during recruitment. All participants took part in individual testing sessions lasting around 30 minutes, with a researcher who was a native speaker of Italian. All recruitment notices, information sheets, consent forms and test materials were provided in Italian.

At the beginning of the session, each participant was given a social issues questionnaire to complete. The topics are listed in Table 1.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Smoking ban:</td>
<td>Should be banned in all enclosed public places?</td>
</tr>
<tr>
<td>2. Capital punishment:</td>
<td>Should be reintroduced in Australia for serious crimes (e.g., murder)</td>
</tr>
<tr>
<td>3. Legalise marijuana:</td>
<td>Should be legalised for public use?</td>
</tr>
<tr>
<td>4. Abortion:</td>
<td>Should be illegal in Australia/Italy?</td>
</tr>
<tr>
<td>5. Same-sex marriage:</td>
<td>Should homosexual couples be allowed to marry?</td>
</tr>
<tr>
<td>6. Sex offender registry:</td>
<td>Should the government require public identification of paedophiles by placing their name, photograph and information on the internet?</td>
</tr>
<tr>
<td>7. Church vs. the State:</td>
<td>Should the Church intervene in political decision-making?</td>
</tr>
<tr>
<td>8. Drink driving:</td>
<td>Alcohol driving limit should be changed from 0.5 to 0.2?</td>
</tr>
</tbody>
</table>

Participants were asked to rate their level of agreement with each issue presented in their questionnaire (1 = strongly disagree, 7 = strongly agree) and also to rate how strongly they personally felt about each issue (1 = no feelings, 7 = very strong feelings). Based on these responses we selected two topics for each participant – one about which they would be asked to give their true opinion and the other about which they would be asked to lie. Wherever possible, we chose issues for which the participant had reported both a strong opinion on the topic (with an opinion rating of either one or seven) and had also expressed strong feelings (provided a value of seven for personal feelings about the issue). The two topics were then randomly assigned to either the truthful or deceptive condition for each participant.

Following completion of the questionnaire, participants were told by the experimenter that they would shortly participate in a taped interview. Participants were told they would be asked to lie or tell the truth about their opinion on some of the social issues that had been presented to them in the social issues questionnaire.

With regard to the topic designated as the truthful target, participants were instructed to simply give an honest account of their genuinely held opinion, while for the topic designated as the deceptive target participants were instructed to lie so that the interviewer would believe they held the opposite opinion to their genuine beliefs. To provide further motivation to appear genuine during both their accounts, participants were told the interviewer was unaware of which account was truthful or deceptive and that the interview would be
terminated should the interviewer become suspicious of the participant at any point.

The order of the two interview topics was counterbalanced with half the participants commencing with a truthful account and the other half commencing with a deceptive account. Once taping of the participants had commenced, the interviewer raised one of the two issues with participants and asked a series of questions about the issue: (i) What is your opinion about this issue?; (ii) Can you tell me why you hold that opinion?; (iii) Is this really your true opinion?; (iv) Are you lying to me now?; and (v) You have told me your opinion, but others might hold the opposite view. Can you tell me what you think might lead them to hold an opposite opinion to yours? At the conclusion of each testing session the participants were debriefed and thanked for their participation.

Amongst the participants, the mean absolute difference of opinion rating from the midpoint of 4 (i.e., mean absolute strength of agreement or disagreement) was 2.85 (SD = .37) for the truthful target topics and 2.80 (SD = .61) for the deceptive target topics. The mean rating of feelings was 6.70 (SD = 2.44) for truthful target topics and 6.20 (SD = 1.15) for deceptive target topics. Paired samples t-tests revealed no significant differences between issues selected for truthful and untruthful accounts in terms of participants’ opinions (t(19) = .295, p = .772, two-tailed) or feelings (t(19) = 1.697, p = .106, two-tailed).

### 2.3 Data preparation and analysis

All of the recordings for each condition were transcribed by a native speaker of Italian. Each account was transcribed verbatim, including interjections such as ‘um’ and ‘em’, which were subsequently included as part of the total word count for each account. Each of the transcripts was tagged for grammatical category using TreeTagger [22], software designed for the annotation of corpora with part-of-speech tags. Accuracy rates for tagging of Italian corpora have been reported at 92% [23]. We chose not to conduct a manual parse of the data, in keeping with fully automated systems of detection that must operate independent of the human observer.

Pronoun counts included all types of stand-alone pronoun: personal (e.g., io, tu, egli, noi, voi), possessive (e.g., mio, tuo, suo, loro, proprio), reflexive (e.g., mi, ti, ci, si, te, ne, lo, la, gli), relative (e.g., che, cui, quale), demonstrative (e.g., questo, quello, costui, ciò), indefinite (e.g., chiunque, ognuno, molto) and interrogative (e.g., che, chi, quanto).

### 3. Results

The participants produced an average of 202 (SD = 77.71) words in the truthful speech condition and an average of 175 (SD = 80.73) words in the deceptive speech condition. Paired samples t-tests revealed no significant differences between the groups for average word count (t(19) = 1.227, p = .235, two-tailed).

Participants showed a decrease in their use of pronouns during their deceptive (M = 10.8%, SD = 3.3%) compared to their truthful speech (M = 12.7%, SD = 2.7%), with a mean difference between conditions of seven pronouns per narrative. The effect size was large (d = .70) and a paired t-test showed that the difference between conditions was significant (t(19) = 2.309, p = .032, two-tailed).

A post-hoc examination was made of the independent frequencies of personal, possessive, reflexive, relative, demonstrative and indefinite pronouns. Interrogative pronouns were excluded from the analyses as only two tokens were used across all 40 narratives. A Bonferroni adjusted alpha level of .0083 (.05/6) was used for each of the pair-wise comparisons. Descriptive statistics for each pronoun type, measured as a percentage of total words in each narrative are provided in Table 2.

Table 2. Means and standard deviations (in brackets) for each pronoun type.

<table>
<thead>
<tr>
<th>Pronoun type</th>
<th>Truth</th>
<th>Lie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal</td>
<td>2.56 (1.38)</td>
<td>2.82 (1.81)</td>
</tr>
<tr>
<td>Possessive</td>
<td>.98 (1.06)</td>
<td>.57 (.60)</td>
</tr>
<tr>
<td>Reflexive</td>
<td>1.34 (1.10)</td>
<td>.73 (.79)</td>
</tr>
<tr>
<td>Relative</td>
<td>4.05 (2.0)</td>
<td>3.96 (1.44)</td>
</tr>
<tr>
<td>Demonstrative</td>
<td>2.48 (1.40)</td>
<td>2.07 (1.01)</td>
</tr>
<tr>
<td>Indefinite</td>
<td>1.03 (1.14)</td>
<td>1.04 (.99)</td>
</tr>
</tbody>
</table>

The average number of pronouns in each type was numerically lower when participants were lying than when they were telling the truth, except in the case of personal and indefinite pronouns, which were marginally higher during deception. However, paired sample two-tailed t-tests showed the differences between conditions for each pronoun class were not significant with t(19) = .693 (p = .497) for personal pronouns, t(19) = 1.519 (p = .145) for possessive pronouns, t(19) = 1.975 (p = .097) for reflexive pronouns, t(19) = 1.46 (p = .886), for relative pronouns, t(19) = 1.010 (p = .325) for demonstrative pronouns and lastly, t(19) = .038 (p = .970) for indefinite pronouns. A discriminant analysis with veracity as the classifying variable revealed that, on the basis of overall pronoun frequency (collapsing across pronoun types), messages could accurately be classified as deceptive or truthful 68% of the time ($\chi^2 (1, n = 40) = 3.87, p < .05$, eigenvalue = .11, Wilks’ Lambda = .90). Cross-validation produced an accuracy rate of 65%.

### 4. Discussion

In a unique contribution to the literature, this study investigated the utility of a linguistic measure of deception in speakers of a language other than English. The aim of the study was to examine the discriminative ability of pronoun use - a variable previously associated with deceptive behaviour in English-speaking participants. It was hypothesised that, as in English, deceptive Italian speech would contain fewer pronouns than truthful utterances.

Our findings supported this hypothesis, with lower proportional use of pronouns during the Italian speakers’ deceptive speech compared with their truthful speech. Moreover, pronoun use accurately classified messages as truthful or deceptive around 68% of the time, which is at a higher level than that commonly reported for the human lie detector (54%; [24]). This novel finding substantiates the status of pronouns as a cross-linguistic variable of interest in the identification of deception. Our post-hoc analyses revealed that the individual pronoun types failed to discriminate between truthful and deceptive utterances even though they reached significance when combined. Hence this study also highlights the methodological value in considering pronouns as an entire category, rather than isolating one or two types of pronoun (such as personal pronouns) for investigation.
The present study was not designed to test probable explanations for a decrease in pronoun use during deception. Future studies may consider manipulating cognitive effort, ambiguity versus specificity, and anxiety (to the extent that it is ethically possible to do so) in order to observe any subsequent effects on pronoun use during deceptive compared to truthful speech. Other grammatical categories, such as noun, verb, adjective and adverb use may also be considered. Additionally, other linguistic measures such as acoustic duration or pitch (of pronouns and other target variables) would be worthwhile to examine, as would the ways in which such variables change during text and computer-mediated communications, compared to speech.

The present study has some limitations. A larger sample size, with a corresponding increase in power, may have produced significant differences within each of the individual categories of pronoun type. Additionally, many participants also spoke English (they were native speakers of Italian living in Australia) and it is possible that their pattern of pronoun usage has been influenced by their exposure to English. We leave these issues to future research. Interestingly, examining deception in a null subject language, such as Italian, may have advantaged the utility of pronoun use in the identification of deceptive compared to truthful speech. For English-speakers, pronoun use primarily involves one task – choosing the correct pronoun. In contrast, for speakers of null-subject languages such as Italian, pronoun use involves an additional task – whether to use a pronoun at all [25].

Future studies might also consider linguistic indicators of deception in the context of bilingualism (most of the participants in the current study spoke two or more languages including English). Furthermore, it is interesting to consider whether the utility of pronoun use (and other linguistic markers of deception) extends to non-Western cultures and languages, such as Japanese which is renowned for using fewer pronouns in general than Indo-European languages.

5. Conclusions
The utility of linguistic markers of deception in English has been widely documented. The present study extended examination of the discriminative utility of one particular linguistic variable, pronoun use, to Italian. The results lend support to the contention that indicators of deception that have emerged in English-speaking contexts may also be valuable clues to deception in other languages.

6. References