対話中における嘘検出に有効な特徴量比較と質問分析

Comparison of Effective Features and Analysis of Questions Towards Dialogue-based Deception Detection

角森唯子[1*]　グラム・ニュービッグ[1]　サクリアニ・サクティ[1]　戸田智基[1]　中村哲[1]
Yuiko Tsunomori[1]　Graham Neubig[1]　Sakriani Sakti[1]　Tomoki Toda[1]　Satoshi Nakamura[1]

[1] 奈良先端科学技術大学院大学
[1] Nara Institute of Science and Technology

Abstract: When humans attempt to detect deception, they perform two actions: looking for telltale signs of deception, and asking questions to attempt to unveil a deceptive conversational partner. There has been significant prior work on automatic deception detection that attempts to learn signs of deception. On the other hand, we focus on the second action, envisioning a dialogue systems that asks questions to attempt to catch a potential liar. In this paper, we describe the results of an initial analysis towards this goal, attempting to make clear which questions make the features of deception more salient. In order to do so, we collect a deceptive corpus in Japanese, our target language, perform an analysis of this corpus comparing with a similar English corpus, and perform an analysis of what kinds of questions result in a higher deception detection accuracy.

1　はじめに

対話中の嘘を見分けることは決して容易ではなく，刑事などは嘘を見抜く高度なテクニックを身につけている [4]．具体的には，嘘の特徴を見分けることと，嘘の特徴がを露呈しやすいように質問を行うことが挙げられる [14]．

近年では，機械学習を用いた嘘の自動検出についての研究が行われており，一定の成果が得られている [5]．文献 [5] では，英語話者によって収録された嘘を含むコーパス (以下， CSC コーパス) に対して，音響特徴量と言語特徴量を用いて嘘の自動検出を行った結果，チャンスレート（60.2%）を有意に上回る分類精度（66.4%）が得られることが確認されている．一方で，言語間，文化間によって嘘の言語的な特徴に差異があることも指摘されている [10]．

これまでの研究は，すでに終了した対話を対象にして嘘の検出を行っている．人間の質問者に置き換えると，これは「嘘の特徴を見分ける」ことに等しい．しかしながら，これは嘘を検出するテクニックの 1 つにすぎず，嘘の特徴が露呈しやすい質問を行うことも同様に必要不可欠である．しかしながら，嘘の自動検出に関する先行研究では，相手に質問をすることは扱われていない．

*連絡先：奈良先端科学技術大学院大学　情報科学研究科
〒 630-0192 奈良県生駒市高山町 8916-5
E-mail: tsunomori.yuiko.tq1@is.naist.jp

本研究では，嘘の特徴を見つけるだけでなく，嘘の特徴が露呈しやすい質問を行い嘘を検出する対話システムの構築を目標とする．本稿では，嘘検出に有効な特徴量の種類だけでなく，嘘の特徴を露呈しやすくさせる質問の種類を確認することによって，目標への第一歩とする．これらを明白にすることによって，嘘の特徴を効果的に誘発する質問に焦点を当てた対話システムの構築が可能となる．

特に，2 点を中心に報告を行う．1 つ目は，我々の対象言語が日本語であることから，英語の嘘を含むコーパスと同じ条件で日本語のコーパスを収集する．このコーパスを用いて嘘の検出実験を行い，嘘の検出に有効であるとされた言語・音響特徴量の比較を日本語と英語で行う．2 つ目は，質問者による質問の種類を分析する．具体的には，分類が容易な発話と難しい発話を誘発する質問それぞれの対話行為について調査する．

2　日本語の嘘を含むコーパス収集とラベル付与

嘘の検出を行うために嘘を含むコーパスが必要であり，いくつかのコーパスが先行研究で作成されている．CSC コーパス [5] は，対象者が質問者に嘘をつくように促されたインタビューを収録し，騙しに成功した場合の報酬によって動機づけされている．インタビューは英語で行われ，25〜50 分の合計 22 のインタビューが収

録されている．その他にも，Idiap Wolf Corpus [6] がある．これは，ロールプレイングゲームに参加する被験者の会話を収録した視覚情報を含む多人数会話コーパスであり，使用言語は英語である．

英語でコーパスが豊富な一方で，他言語のコーパスは数少ない．日本語のものとしては，インディアンポーカーコーパス [9] がある．これは，インディアンポーカーゲームの会話を収録した視覚情報を含む多人数会話コーパスである．しかしながら，本研究は一対一の対話システムを構築することを目標にしているため，多人数話者によるコーパスを用いるのは適切ではない．本節では，言語と文化間で言葉で嘘検出の研究を比較し，日本語の嘘検出対話システムを構築するために，日本語の嘘を含むコーパスを収集する．

2.1 コーパス収集

CSC コーパス [5] と同様の収録設定の下，日本語で対話収録を行う．嘘が誘発されやすい場面の一例として，面接 [3] における質問者と対象者の 2 名の対話を想定する．収録の手順を以下に示す．

1. 対象者が 6 項目（政治，音楽，地理，食べ物，インタラクティブ，サバイバル）に対する，ある“ 目標プロフィール ”に適合するかどうかを調べるための実験であると対象者に伝える．
2. 対象者に 6 項目のテストを収録前に受けてもらう．
3. 実験者は，2 項目が適合，4 項目が不適合となるように結果を操作し，その結果を対象者に伝える．
4. この実験の本当の目的は“目標プロフィールに適合している ”と主張し，質問者を納得させることができる人物を探すことであり，上手く納得させれば賞金があると対象者に伝える．
5. 対象者は，全項目のテスト結果において“目標プロフィールに適合している ”と，面接で質問者に主張する．質問者は，目標プロフィールやテスト内容についての知識はなく，いかなる質問をしても構わないとする．

20 代の男女計 2 名を質問者とし，20 代の男女計 10 名を対象者として対話収録を行った．収録した総対話数は 10 対話であり，総時間は約 150 分である．対象者の総発話数は 1069 で，SU[1] の総数は 1671 である．この日本語偽言コーパス（JDC) は，研究目的に限り使用することができる [2] ．収録した日本語偽言コーパスの一例を表 1 に示す．

[1]発話を句読点や休止で区切ったスラッシュユニットの略称
[2]http://ahclab.naist.jp/resource/ja-deception/

表 1: 対話例 (I：質問者，P：対象者)

話者	書き起こし文	ラベル
I	音楽に関して，あなたはマッチしていましたか？	
P	はい，マッチしていました．	嘘
I	それはなぜだと思いますか？	
P	えーと，そこそこ答えれたからです．	真実
P	小さい頃からずっとピアノをやっていたので．	嘘

表 2: 音響および言語特徴量と個人性の情報

カテゴリ	説明
一般記述	テスト項目，笑い，雑音，言い間違い
文構造	代名詞，否定，YesNo，終助詞，動詞の原形，遊び言葉，合図句，質問，ポジティブ・ネガティブ単語
パラ言語情報	同意，言いよどみ
F_0	中央値，SU 長に対する中央値の割合
音素継続長	母音，平均，最大
パワー	平均，SU の第 1・最終フレーム
個人性	性別，言いよどみ・合図句の頻度

2.2 ラベル付与

対象者の SU に対してラベル付与を行う方法として“真実 ”と“嘘 ”のボタンを用意し，対象者には対話収録中の SU ごとにどちらかのボタンを押してもらう．なお，一部分でも嘘が含まれる SU は，嘘として定義する．それぞれのボタンが押された時間と書き起こしから，SU 毎にラベル付けを行う．1671 SU のうち真実ラベルが付与された SU は 1401 で，嘘ラベルが付与された SU は 270 である．

3 嘘検出のための特徴量

嘘の検出実験を行うために，嘘の手がかりを示す特徴量を抽出する．本研究は先行研究 [5] に基づき，SU 毎に言語特徴量と音響特徴量に着目する．抽出した特徴量を表 2 に示す．

- **音響特徴量**
  音響特徴量として, 基本周波数 (F_0) とパワー，音素継続長を用いる．F_0 の抽出には Snack Sound Toolkit[12]，音素継続長の抽出には Kaldi[11] を用いる．
- **言語特徴量**
  人手による書き起こし文に対して MeCab[7] で形態素解析を行い，得られた形態素から言語特徴量を抽出する．雑音は，咳やマイクとの接触音など対象者によって生成されたものに限定する．ポジティブ・ネガティブ単語の頻度の抽出には，単語

表 3: 嘘の検出率：音響特徴量 (AP)，言語特徴量 (L)，個人性 (S)

	日本語		英語	
特徴量	Accuracy(%)	F 値 (%)	Accuracy(%)	F 値 (%)
Chance rate	83.8	0.0	71.4	0.0
AP	90.5	60.2	86.8	74.5
L	84.2	7.6	71.4	14.7
AP+S	90.7	61.4	88.1	77.7
L+S	85.2	31.5	76.8	52.9
AP+L	89.9	56.9	86.8	74.6
AP+L+S	90.2	58.1	87.8	77.2
人間	83.0	28.4		

表 4: 被験者ごとの嘘の検出率

被験者	A	B	C	D	E	F	G	H	I	J
Chance rate (%)	93.5	89.4	92.4	76.9	78.7	75.9	64.9	82.9	73.1	83.9
Accuracy (%)	93.2	89.4	93.2	80.2	86.5	88.6	84.7	87.4	73.1	93.7

感情極性対応表 [13] を用いる．英語の先行研究 [5] で用いられた特徴量に加えて，“終助詞の数”を新たに加える．

- **個人性**
  対象者の性質に依存する特徴量を抽出する．対象者の性別と，1 対話中の全 SU 数に対する合図句・言いよどみの頻度を個人特徴量とする．

4 嘘検出の実験的評価

前述した音響特徴量と言語特徴量を合わせた特徴量を用いて，SU に対して嘘の検出を行う．2 値分類するための機械学習手法として，Bagging[1] を用いる．嘘の検出率の評価には，1670 SU を学習用のデータにし，残り 1 SU をテスト用のデータにする一個抜き交差検定を用いる．

4.1 結果と考察

表 3 に，各特徴量を用いた際の嘘の検出率と F 値を示す．“日本語”は本研究における日本語の嘘検出率，“英語”は先行研究 [5] に基づき，CSC コーパスの一部を用いて英語の再現実験（日本語で加えた特徴量は加えていない）を行った際の嘘検出率である．“人間による検出”とは, 被験者と別の人間に同じ対話データを聞いてもらい, 嘘の検出を行ってもらった結果である．なお，検出の際に文脈は考慮しない．

日本語では音響特徴量+個人性を用いた場合が最も精度が高く，言語特徴量のみに比べて嘘検出率が約 7% 高かった．全特徴量を用いた場合でも，音響特徴量+個人性，音響特徴量のみの検出率より低い精度であった．英語でも音響特徴量+個人性を用いたものが最も精度が高く，チャンスレートに比べて約 17% 高かった．人間による検出はチャンスレートとほぼ同等の精度であり，どの特徴量を用いた検出もこれを上回る精度であった．また，日英ともに言語特徴量を用いた場合の検出率はチャンスレートとほぼ同等であった．さらに，言いよどみと合図句の頻度，性別などを追加して検出を行った場合に精度の向上が見られたことから，日英ともに嘘検出には個人性が影響を及ぼすことがわかった．なお，フィッシャーの正確確率検定の 1% 水準を用いて特徴量間の違いを比較したところ，日英ともにチャンスレートと音響特徴量，音響特徴量+個人性，音響特徴量+言語特徴量，音響特徴量+言語特徴量+個人性で p 値が 0.01 以下となり，有意差が見られた．

表 4 に，話者ごとの嘘の検出率を示す．検出率が低いほど，嘘が見破られにくい被験者ということを表している．表 5 に，嘘の高検出率と低検出率の対象者の対話例を示す．SN は雑音，MP は言い間違いのことを指す．質問者がテストの出来について質問した際の対話であり，G は検出率が高かった話者，A は低かった話者である．G は雑音や言い間違いが多く，語尾が上ずっていた．その一方で，A は真実の発話に比べて際立った変化は見られなかった．

表 5: 対話例（G：高検出率，A：低検出率）

対象者	書き起こし文
G	$\langle SN \rangle$はー，そうですね，まぁ$\langle MP \rangle$七八割は答えれたかなっていう位ですね．
A	たぶん大丈夫だと思います．

表 6: 特に有効な特徴量

カテゴリ	英語	日本語
言語	雑音，代名詞，YesNo	動詞の原形
個人性	合図句の頻度	
F_0	中央値	中央値
音素継続長	平均，母音	母音
パワー	第 1・最終フレーム，平均	最終フレーム

4.2 特に有効な特徴量

本章では，嘘の検出に有効な特徴量について日英比較を行う．最良優先探索を用いて，分類率が学習データに対して最大となるように特徴量選択を行った．その結果，特に有効とされた特徴量を表 6 に示す．音響特徴量に関しては F_0 の中央値，母音の平均継続長，パワーの最終フレームが日英ともに有効であることが確認された．これらの特徴量が日英ともに有効であるとされた理由を，以下に述べる．

- **パワーの最終フレーム**
  一般に，感情の変化（不安など）は発話の語尾に現れると言われている．
- **F_0 の中央値**
  嘘をつく際は，声の高さに変化が出る [3].
- **母音の平均継続長**
  嘘を話す場合と真実を話す場合を比較すると，話す速度が異なる [8].

言語特徴量は，日英で有効な特徴量が大きく異なっている．英語では，代名詞と YES・NO の有無が有効であるとされた．一方で，日本語では動詞の原形の有無のみであったことからも，今回用いた言語特徴量はあまり貢献しないと考えられる．

5 嘘の検出に有効な質問分析

本研究の目的は，嘘を検出する対話システムを構築することである．本章では，この対話システムが効果的に嘘の検出を行うために，行うべき質問の種類を分析する．全ての返答を対応する質問のクラスに分類し，それぞれのクラスごとに嘘検出率を計算する．あるクラスにおける嘘検出率が高ければ，その質問の種類が嘘の特徴を効果的に引き出すことができ，容易に嘘を検出できると考えられる．そして，この種類の質問に焦点を当てて対話システムを構築することで，より高精度のシステムの作成が可能であると考える．

5.1 対話行為ごとの分析

まず，質問者の発話の種類（対話行為）が嘘検出率に影響を及ぼすと仮定した．この仮説を検証するために，返答を分類するクラスとして，質問の対話行為を使用する．ISO 国際標準により定められた一般目的機能（general-purpose functions: GPF）を利用し，質問者の発話に GPF タグを付与する（ISO24617-2, 2010）．GPF タグの定義を以下に述べる．

- **CheckQ**
  聞き手によって与えられた命題の真偽について，話し手が確信が持てていない場合に使用．話し手が命題の真偽を確認するため，聞き手に対して情報提供を促す．
- **ChoiceQ**
  話し手が，聞き手が知っていると仮定する命題のリストの中における真の要素を知るために，聞き手に対して情報提供を促す．
- **ProQ**
  話し手が，聞き手が知っていると仮定する命題の真偽を得るために，聞き手に対して情報提供を促す．
- **SetQ**
  話し手が，聞き手が知っていると仮定する集合のどの要素が固有の特性を持つかを知るために，聞き手に対して情報提供を促す．

本稿では，JDC のうち同一質問者で収録された 8 対話に対して，人手で GPF タグの付与を行う．JDC の 10% に対し 2 人のアノテータがタグの付与を行った結果，一致率は約 80% であった．これより，第一アノテータが付与した GPF タグに従う．表 7 に，GPF が付与された質問のクラスごとの嘘検出率を示す．ユーザが嘘をついている際にシステムが正しく検出を行うために，嘘の再現率に着目する．図 1 に嘘の再現率を示す．信頼区間は，有意水準 p<0.05 の Clopper-Pearson 法を用いて求める．最も正しく嘘を検出できているのは，CheckQ のクラスである．CheckQ への返答において，対象者は再度同じ話をする傾向にあった．対象者の自信を揺るがして嘘を露呈させるために，再度同じ話を

表 7: クラスごとの嘘検出の詳細

	再現率 (%)		Rate(%)	
クラス	真実	嘘	Accuracy	Chance rate
CheckQ	99.4	83.3	95.3	77.7
ChoiceQ	100.0	80.0	96.4	78.6
ProQ	100.0	69.1	91.5	66.7
SetQ	99.5	73.7	94.0	75.8

させることが有効であると [8] で述べられていることからも，CheckQ が嘘の検出に有効であることがわかる．その一方で，最も検出率が低かったのは ProQ である．ProQ はイエス・ノーで返答することから，対象者への心理的抑圧が少なく，嘘の特徴が露呈しにくかったと考えられる．

さらに，表 8 に，各クラスの返答の SU あたりの単語数（平均 SU 長）を示す．CheckQ のクラスの返答の平均 SU が最も短かった．極端に短い発話において嘘が露呈しやすいと [8] で述べられていることからも，CheckQ は嘘検出に有効な質問であると言える．

5.2 質問長ごとの分析

嘘を見破るために，嘘の特徴を露呈しやすくさせる質問を行うことは重要である．話の細部について質問し，相手の自信を揺るがすことで，嘘の特徴を露呈させやすいといわれている [8]．特に，CheckQ のような再度同じ話をさせる質問に対して，対象者は作り話の細かい点にまで答えざるを得ない場合は多い．このような場合において，質問者が質問を行っている時間に対象者が作り話を考えるという点から，質問長は重要な要素であると考える．ここで，質問長が短いと対象者が嘘を考える時間が短くなるために嘘が露呈しやすくなり，質問長が長いと巧妙な嘘をつくことができると仮定する．

これを調べるために，質問長ごとの嘘の検出率について検証する．図 2 に，質問の SU 長ごとの嘘の再現率を示す．信頼区間は，有意水準 p<0.05 の Clopper-Pearson 法を用いて求める．1 ～ 10 は，SU 長が 1 ～ 10 単語の質問に対する返答を分類した際の嘘の再現率を指す．SU 長が 1 ～ 10 単語が最も再現率が高いことから，SU 長が短い質問が嘘の検出には有効であるとわかる．SU 長が短い質問を行い対話の展開を早く行うことで，対象者の嘘が露呈しやすいと考えられる．今後は，SU 長ごとの返答内容の分析を行う．

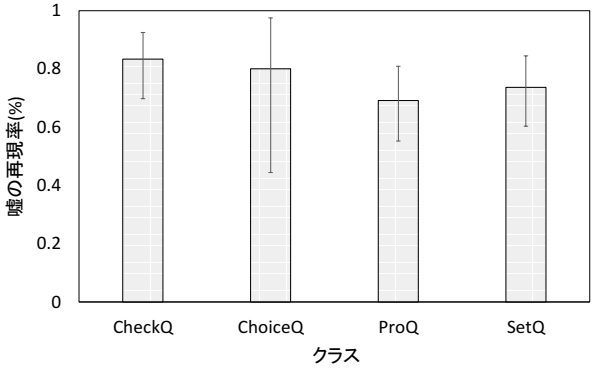

図 1: クラスごとの嘘の再現率

表 8: クラスごとの対象者（返答）の平均 SU 長

	クラス				
	CheckQ	ChoiceQ	ProQ	SetQ	平均
SU 長	6.4	12.2	11.8	18.1	12.3

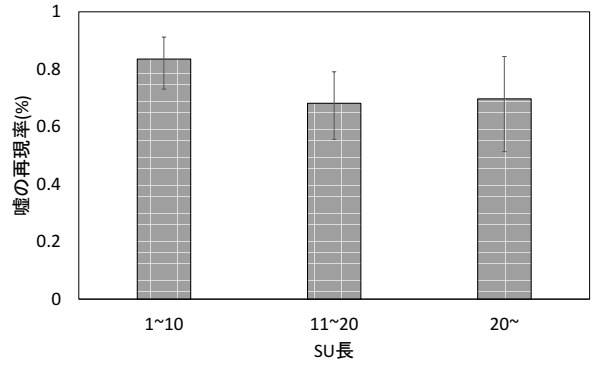

図 2: 質問の SU 長ごとの嘘の再現率

6 まとめ

本稿では，JDC の収集と嘘の検出実験ののち，有効な特徴量の日英間比較を行った．先行研究 [5] で有効とされた特徴量を用いて分類を行った結果，日本語でもほぼ同等の精度が確認された．特徴量に関して属性選択を行い，音響特徴量は日英ともに有効であることを確認した．さらに，質問者の質問の種類と，嘘検出の容易性についての関係の分析を行った．CheckQ が効果的に嘘の特徴を引き出すことができ，嘘の検出に最も有効であることを確認した．また，短い SU 長の質問を行い対話の展開を早くすることで，相手の嘘が露呈しやすいという結果を得た．

今後は，質問者の音響特徴量を加え，声質という点

からも嘘の検出に有効な質問の分析を行う．さらに，効果的な質問の分析に基づく発話を行う嘘検出対話システムの構築を目指す．

参考文献

[1] L. Breiman. Bagging predictors. *Machine learning*, Vol. 24, No. 2, pp. 123–140, 1996.

[2] B. M. DePaulo, D. A. Kashy, S. E. Kirkendol, M. M. Wyer, and J. A. Epstein. Lying in everyday life. *Journal of personality and social psychology*, Vol. 70, No. 5, p. 979, 1996.

[3] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. Cues to deception. *Psychological bulletin*, Vol. 129, No. 1, p. 74, 2003.

[4] P. Ekman. *TELLING LIES*. W. W. Norton & Company, 1985.

[5] J. B. Hirschberg, S. Benus, J. M. Brenier, F. Enos, S. Friedman, S. Gilman, C. Girand, M. Graciarena, A. Kathol, L. Michaelis, et al. Distinguishing deceptive from non-deceptive speech. *Proc. Eurospeech*, 2005.

[6] H. Hung and G. Chittaranjan. The IDIAP wolf corpus: exploring group behaviour in a competitive role-playing game. In *Proc. The international conference on Multimedia*, pp. 879–882. ACM, 2010.

[7] T. Kudo, K. Yamamoto, and Y. Matsumoto. Applying conditional random fields to japanese morphological analysis. *Proc. EMNLP*, Vol. 4, pp. 230–237, 2004.

[8] P. Meyer. *LIE SPOTTING*. Griffin, 2011.

[9] Y. Ohmoto, K. Ueda, and T. Ohno. A method to detect lies in free communication using diverse nonverbal information: Towards an attentive agent. In *Active Media Technology*, pp. 42–53. Springer, 2009.

[10] V. Pérez-Rosas and R. Mihalcea. Cross-cultural deception detection. *Proc. ACL*, pp. 440–445, 2014.

[11] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely. The kaldi speech recognition toolkit. *Proc. ASRU*, 2011.

[12] K. Sjolander. Tcl/tk snack toolkit. 2004. `http://www.speech.kth.se/snack/`.

[13] H. Takamura, T. Inui, and M. Okumura. Extracting semantic orientations of words using spin model. *Proc. ACL*, pp. 133–140, 2005.

[14] A. Vrij, P. A. Granhag, S. Mann, and S. Leal. Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science*, Vol. 20, No. 1, pp. 28–32, 2011.