

間違いを含んだ会話システム

7R-5

平松 薫* 山本 吉伸† 安西 祐一郎‡

慶應義塾大学§

1 はじめに

対話処理に関する研究のテーマは、意味や談話に関する深い解析へと移り大きな成果をあげているが、まだ自然言語処理自体の問題は多い [1]。対話において意味や談話などを解析し、動的な主題の変化に対応するためには、計算コストは大きいと予想される。また、「相手が計算機である」という認知によって、相手を人間と認知した場合とは違う反応が現れる可能性がある。対話中の文法的な誤りでも人間と認知されている場合なら問題ないが、計算機と認知されている場合には即座にブレークダウンを起こす可能性がある [2]。

2 インタフェースの支援

そこで、本稿では、ネットワークを介してのテキストのみによる対話システムを取り上げ、人間同士の対話と自然言語対話システムとの対話を比べ、自然言語処理から離れた、インタフェース的立場で自然さを補う要素をとって、次のものを検討する。

- (1) 文単位の出力を文字単位の出力に変換し、人間のキータイプのスピードとミスタイプのシミュレートする
- (2) 会話中でキーワードを発見しだい、その場であいづちを生成する
(例：“しづやにあるおみせ”の“しづや”からあいづちとして“はい”を生成する)

(1) はフィルタプログラムによって実装した。
(2) は会話プログラムによって実装し、キーワードとなる単語はあらかじめ対話の主題ごとに登録した。この二つの要素について、会話における効果を実験から検討する。

現在の計算システムでは、処理速度の向上にともない数行の文章ならば、一瞬で表示できる。しかし、人間のキー入力速度は、熟練者であっても4~5(文字/秒)程度であり、ミスタイプをする可能もある。画面上でのネットワーク対話プログラム“phone”では、打ったキーがそのまま相手の画面に反映されるため、プログラムによる文章出力と

人間のキー入力は大きなスピードの差を生む。そのキータイプ速度が自然な会話を妨げる原因の一つになると考えられる。そこで、対話に(1)が与える影響を検討する。

一方、人間同士の会話であれば、相手の発話中に(割り込んで)あいづちを場合がある。あいづちはそのタイミングによって、相手に対する深い理解や、軽い受け答えを示し、聞き手の理解や同意などの状態を表す [3]。また、発話の順番やリズムに対話の基本的な情報が含まれていること [5] や、会話の“間”などにも、一定の法則が存在する [6] ことがわかっている。(2)では、あいづちの対話に与える効果を検討する。

3 実験と評価

Loebner 賞大会 [4] での評価方法に沿って実験、評価を行なった。被験者にプログラムと実際に会話させ、その評価によって要素として取り上げた上記の(1)(2)について検討する。

3.1 被験者

phone プログラムの使用に支障のないキータイプができる大学生12名であった。

3.2 実験計画

ネットワークでつながった計算機を離れた場所に2台用意し、phone プログラムを通して実験を行なった。被験者は、「phone で会話をしながら、相手が人間かプログラムか当てて下さい。」という課題のもとで会話を行わせた。対話の相手はプログラム(4条件、P1~4)と人間(2条件 Ha Hb)の6条件を用意した。各条件ごとに会話に異なる主題を用意し、その主題から離れた内容については相手が答えないこととした。プログラム相手の場合は表1に示す条件で実験を行なった。相手が人間の Ha Hb の条件は実験スタッフが対話相手となった。

各パターンが終了後、相手がどちらか(人間、プログラム)、判断の根拠(記述)、相手に対して持った印象(5段階評価、14項目)を調べ、全パターン終了後、会話の相手を“人間らしさ”で順位付けをさせた。

*Kaoru Hiramatsu

†Yoshinobu Yamamoto

‡Yuichiro Anzai

§Keio University

表 1: 対話プログラムのパターン

プログラム	キータイプ (1)	割り込み (2)
P1	無し	無し
P2	有り	無し
P3	無し	有り
P4	有り	有り

3.3 結果、考察

被験者のつけた順位とその判断時間の関係を表したのが、図 1である。会話において長時間の沈黙はないので、時間が判断の難しさを表すことになる。実際に順位と判断時間には相関があり、出力が人間らしくなると人間かプログラムかの判断が難しくなることを示している。

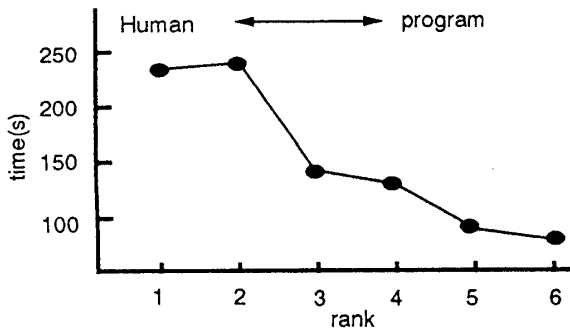


図 1: 判断時間

図 2は各パターンごとの順位の平均である。被験者の判断は Friedman の検定 (残差平方和 $S=1520$, 一致性の係数 $W=0.603$, $F_0 = 16.70$, $\phi_1 = 4.83$, $\phi_2 = 53.17$, 危険率 1%) により、一致してしていることが確かめられた。人間が相手の場合とプログラムが相手の場合では差は大きい。要素 (1)(2) を加えると順位が人間相手の場合に近くなることわかる。

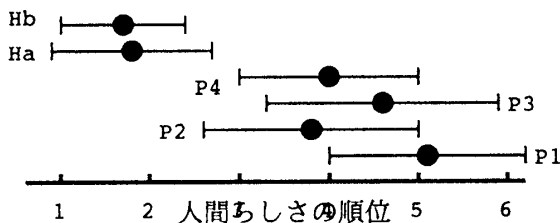


図 2: 各パターンの順位

5段階評価の 14 項目について相関行列を求め、相関の値によってグループ分けをした。(評価項目: 親しみやすい、(反応が)遅い、ふざけている、(会話が)自然な、子供っぽい、(相手) 個性的、冷たい、(会話が)活発な、感情的、(相手) 好意的、

会話がつながらない、打ち間違いが多い、(文章が) 作爲的、出力が速い)

(a) 会話へのプラスの印象の項目

親しみやすい、反応が遅い、会話が活発 (会話が) 自然な、(相手が) 好意的、打ち間違いが多い

(b) 会話へのマイナスの印象の項目

冷たい、(文章が) 作爲的、出力が速い 会話がつながらない、(相手) 個性的

また、要素に対する評価と対話に対する印象の項目のうち相関の大きかったものは、次の通りである。

正の相関 冷たい-出力が速い (0.931)

(会話が) 自然な-打ち間違いが多い (0.705)

親しみやすい-打ち間違いが多い (0.785)

負の相関 (会話が) 自然な-出力が速い (-0.771)

冷たい-打ち間違いが多い (-0.846)

被験者の主観評価において、ミスタイプに対する評価は会話の自然さと正の相関があり、phone を通した対話において要素 (1) の速度とミスタイプのシミュレートが有効であることを示している。

要素として検討した割り込み発話 (2) は、判断基準の記述においても、タイミングがおかしいなどの指摘があり、キーワードからだけのあいづちでは、会話の自然さを高めることができなかつた。あいづちなど割り込み発話において自然さを補うには、正確なタイミングとキーワードの段階よりも上の言語処理とのリンクが必要と考えられる。

4 まとめ

対話システムの要素 (1)(2) について、有効性を検討した。キータイプシミュレート (1) については、その使用により対話システムの“自然さ”を補えることを主観的評価で、その有効性を示した。あいづちに関する (2) は、キーワードからあいづちを生成するシステムでは有効性は示せなかつた。

参考文献

- [1] 徳永 健伸, “自然言語生成への招待”, 情報処理, Vol. 33, No. 7, 1992
- [2] 山本 吉伸, “コミュニケーションの面白さ -しりとりはなぜおもしろいか-”, ヒューマンインタフェース, 47-13, pp. 426-433, 1993
- [3] 小坂 直敏, “あいづちを中心とした会話音声の呼応関係の分析”, NLC 87-23, 1987
- [4] 大沢 英一, “コンピュータは人間をだませるか -チューリングテストをめぐって-”, 情報処理, Vol. 34, No. 3, pp 280-283, 1993
- [5] Satoru Hayamizu, “Role of rhythm and tempo in spoken dialogue”, Proceedings ISSD, pp 177-180, 1993
- [6] 中村 敏枝, ““間”に関する心理学的研究”, 信学技法, HC92-71, 1993