

FGS：言葉とジェスチャを用いた情報表現に関する研究

FGS: A study of information representation with words and gestures

藤本健太^{1*} 今井倫太¹

¹ 慶應義塾大学

¹ Keio University

Abstract: 本研究では、人型のロボットを用いて、ユーザの位置情報によってコンテンツ内容に変化を加えるシステム FGS を提案する。そのために、ユーザがコンテンツが話題にした現場にいたかいなかったか、ジェスチャを主として用いるか否かで、それぞれコンテンツのモデルを作成し、各モデル間に互換性があるかどうかを検討した。結果として、より多くの情報を含む、ユーザが現場にいないことを想定した文章からの変換がもっとも容易だということが分かった。

1 はじめに

本稿では、携帯可能な小型のロボットである BlogRobot(図1)を用いて、ユーザの現在位置に応じて発話およびジェスチャを変化させるモデル FGS に関して検討を行う。

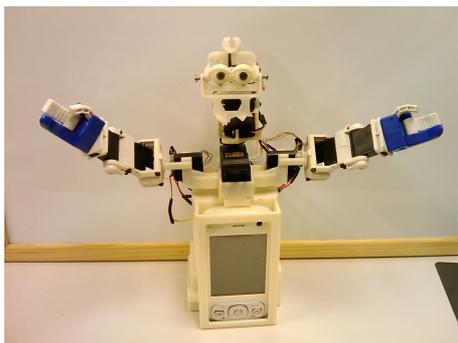


図 1: BlogRobot

GPS 機能付きの携帯電話が普及し、ユーザの現在位置に応じた様々なサービスが受けられるようになった。しかし、ユーザの現在位置に応じて、どのようにコンテンツ情報を提示すれば効果的にユーザに伝達できるかというのは十分な議論が必要だと思える。

株式会社トランスメディアが行っている mapii は、GPS とブログの連動サービスである。[1] 携帯端末上に簡易マップを表示し、マップに友人の現在地の位置情報や口コミ情報の表示が可能になっている。SekaiCamera[2] や実世界透視ケータイ [3] では、GPS に加えて内部の加速度センサや電子コンパスを用いてディスプレイ上にカメラで見える範囲の情報を提示することを可能にした。また、我々は以前の研究で、ディスプレイを介し

た web ブラウザを用いた閲覧ではなく、BlogRobot による発話・ジェスチャの表現をもちいてユーザに記事を提示する手法を提案した。[4] その際、ブログが話題にしている場所(以下、対象地点)をより直感的にユーザに理解させるため、対象地点の方向を BlogRobot が指差す手法を取り入れた。

以上のように位置情報を用いたサービスは、コンテンツの表示方法に様々な工夫がなされている。しかし、コンテンツ内容に関しては、読み手の現在位置を考慮に入れていない。例えば、書き手が「
という洋服屋を利用した」という位置情報付きのコンテンツがあった場合、読み手が、お店の目の前にいるならば「目の前の洋服屋を利用した」と、より実世界を意識させたコンテンツ情報にしたほうが記事の内容の理解が早まると思える。提示手法のみを検討するのではなく、ユーザの状況に合わせて情報本体にも変更を加えることが必要である。

本稿では、BlogRobot を用いて、ユーザの位置情報によってコンテンツ内容に変化を加えるシステム FGS を提案する。FGS を実現するために下記の内容について検討を行う。

- ユーザが現場にいることを想定した文章から、現場にいないユーザ用の文章に変換可能か
- ユーザが現場にいないことを想定した文章から、現場にいるユーザ用の文章に変換可能か
- ユーザが現場にいること想定した文章から、ジェスチャメインの文章へ変換可能か

本稿の構成は以下のとおりである。2 章では、本稿が扱う課題について述べる。3 章では、読み手ユーザの環境に応じた会話モデルのルールを設定し、実例を示す。4 章では、提案である FGS の動作イメージについて述べる。5 章で考察を行い、6 章で結論を述べる。

*連絡先：慶應義塾大学理工学部情報工学科 安西・今井研究室
〒 223-0061 神奈川県横浜市港北区日吉 3-14-1
E-mail: fujiken@ayu.ics.keio.ac.jp

2 記事の再生とジェスチャ

本章では、BlogRobot の概略に関して説明し、本稿で扱う課題の詳細を述べる。

2.1 BlogRobot

図1の BlogRobot は、ロボットによる音声発話とジェスチャを用いて、ブログ記事をユーザに提示するロボットである。ロボットという実体を持つことで、実世界的な表現（例えば、実際の建物を指差す）やジェスチャによる非言語的表現を可能とする。本章では、BlogRobot の仕様について説明する。

2.1.1 音声発話・ジェスチャ制御

BlogRobot は PDA を搭載しており、PDA 上で音声合成による音声発話を行う。また、PDA からの命令によってサーボモータが動作する。複数のサーボモータを制御することで各種ジェスチャを可能にしている。今回、ジェスチャは大きく分けて以下の4種類を用いる。

- 指示：物体の方向を指差す
- ストローク：物体の動きを表現する「まっすぐ」「直角に」など
- エンブレム：物体の形態を表現する「丸い」「四角い」など
- ビート：腕を少し上げることで、発話の強調を行ったり、調子を整えたりする。

2.1.2 記事の再生手順

BlogRobot がブログの記事を再生する手順は図3の通りである。まず、専用のオーサリングツールを用いてブログ記事を作成する。[5] このオーサリングツールでは、通常表示する文章とは別に、書き手が任意に動作アイコンを挿入する。作成されたブログの記事の例を図2に示す。

この記事が専用のサーバを介することで、通常表示の文章は発話命令に、動作アイコンはジェスチャ命令となり、動作命令テーブル(図4)が生成される。動作命令テーブルには、音声発話命令およびジェスチャ命令によって構成されており、PDA 上のプログラムによって上から順番に実行される。



図 2: BlogRobot 専用のオーサリングツールで作成された記事

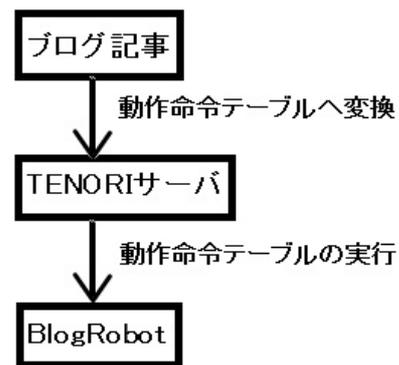


図 3: BlogRobot による記事の再生手順

2.1.3 対象への指示機能

ブログの記事上に GPS 情報を含めることで、BlogRobot は記述された GPS 情報の地点を指差しする機能を有する。ブログ記事中の GPS 情報は、GPS 情報を含んだ写真データを添付することで簡単に挿入できるようになっている。記事中の GPS 情報と、BlogRobot に搭載された PDA が取得する自身の GPS 情報と方位情報を用いることで対象への指差しが可能となっている。

指示機能を用いた例を図5に示す。先頭の GPSAction という命令は、対象を指差すためのジェスチャ命令であり、ブラウザ上では表示されない。また、簡単のため、そのほかのジェスチャ命令は省略しているが、書き手が任意のタイミングで挿入することが可能である。例では、文章を音声で発話する前に GPSAction 命令によって、「どこそこを見て」という対象(タチバナ)への

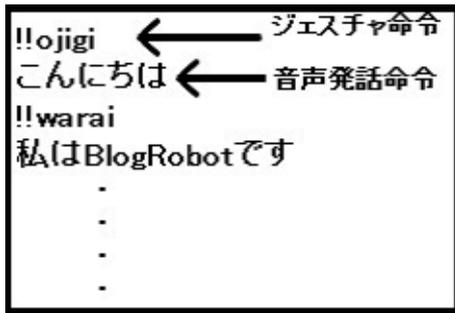


図 4: 動作命令テーブル

方向指示を指差しジェスチャを用いて行う。指差しジェスチャによって、ユーザはブログ記事が話題に出した場所を地図などを用いるよりも早く理解することが出来る。

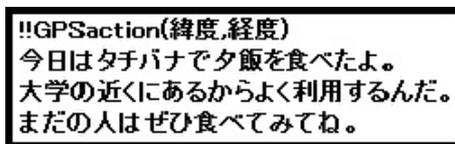


図 5: 方向指示機能を含む記事の例

2.2 複数環境下における記事再生の課題

以前までの BlogRobot では、ユーザが記事の現場 (例えばタッチバナ) にいることが前提となっていた。これは、ユーザの現在位置をリアルタイムに監視し、記事内に埋め込まれた GPS 情報との距離が一定値以下になったら再生するという、ユーザの現在位置によって記事の提示を行っていたためである。しかし、将来的にユーザの現在位置以外の方法で記事を提示することになった場合、必ずしもユーザが現場にいるとは限らない。この際、ユーザが現場にいるかいないかによって、BlogRobot の振る舞いを変える必要が生じる。例えば、指示機能に関して、ユーザが現場にいるときは「あっち」「こっち」など指示語を用いることが出来たが、ユーザが現場にいない場合は指示語は使えず、固有名詞や物体の動きや形態を示すストロークジェスチャやエンブレムジェスチャなどを用いて事細かに対象を説明する必要がある。つまり、書き手は 1 つの記事に対して 2 通りの記述を用意しなければならず、記事作成に負荷がかかる。

3 ブログ記事の自動生成

3.1 モデル間の変換

読み手の環境を考慮に入れることで、書き手が記事を作成する負荷が増加する。書き手の負荷を軽減するため、1 つの記事から読み手の環境情報を考慮した複数記事の自動生成手法を検討する。自動生成に関して検討する前に、読み手ユーザの現在位置に応じた以下のモデルを定義する。

- L 型モデル (L) : ユーザが現場にいないことを想定して書かれるモデル
- N 型モデル (N) : ユーザが現場にいることを想定して書かれるモデル
- S 型モデル (S) : ユーザが現場にいることを想定し、かつジェスチャをメインで用いるモデル

3 つ目のモデル S に関しては、音声発話を削り、ジェスチャを主として使うモデルである。例えば、複雑な動きをわざわざ口で説明するよりもジェスチャで表現したほうが早く正確に伝えられることが多い。ユーザへより円滑に情報が伝達できるモデルとして S モデルを考慮に入れる。

各モデル間の変換を自動で行うシステムを検討することで、書き手の負荷を加えずに、読み手の状況に即した記事の再生を実現できる。下記のモデル変換の検討を行う。

- N → L 変換 : ユーザが現場にいることを想定した文章から、現場にいないユーザ用の文章に変換可能か
- L → N 変換 : ユーザが現場にいないことを想定した文章から、現場にいるユーザ用の文章に変換可能か
- N → S 変換 : ユーザが現場にいることを想定した文章から、ジェスチャメインの文章へ変換可能か

3.2 各モデルのルール

各モデル間の変換を行うために、上の項で挙げた各モデルに関してルールを設定する。

3.2.1 文章に関するルール

ユーザが現場にいるかいないかによって以下のことが想定される。

- 読み手が現場にいない状況では指示語を使えない

- 読み手が現場にいない状況では、説明のために文章による記述を多く用いる必要がある

以上を踏まえた上で、BlogRobot の発話に使われる文章に関しては次のルールを設定する。

- 物体指示
 - L 物体の特徴による詳細を記述（赤い丸いりんご，顔だけのロボット）
 - N 物体との位置関係による詳細を記述（机の上にあるバナナ）
 - S 指示語（あれは，これは）
- 場所の参照
 - L 固有の名称（慶應義塾大学，定職屋タチバナ）
 - N 指示語と名詞の組み合わせ（この部屋，左の場所）
 - S 指示語（ここ，あそこ）
- 修飾句
 - L 複数付加する（大きな白い花）
 - N 1つ付加する（大きな花）
 - S 付加しない（花）

3.2.2 ジェスチャに関するルール

物体や場所の指示は、指示ジェスチャを用いることが有効である。しかし、読み手が現場にいない状況での指示語は使えないので、Lモデルの場合は指示ジェスチャは使えない。また、Sモデルと他のモデルとの差別化を行うために、物体の様態をあらわすジェスチャ（ストローク、エンブレム）はSモデルのみ利用することにする。ビートジェスチャはどのモデルも使用する。以上を表1にまとめる。

ジェスチャの種類	L	N	S
指示	×		
ストローク	×	×	
エンブレム	×	×	
ビート			

表 1: 各モデルで利用するジェスチャ

3.3 スクリプト

発話文で構成されるスクリプトを作成した。(図6, 図7, 図8)

01: みなさん、はじめましてこんにちは
 02: 今から安西・今井研究室という慶應義塾大学の研究室を紹介するよ。
 03: 部屋番号25の419は、学生たちが主にデスクワークする場所なんだ。
 04: だから、あちこちがパソコンだらけで散らかっているよ。
 05: パソコン以外には、研究で使うユニークなロボットが沢山あるよ。
 06: いくつかロボットを紹介するね。
 07: まず、部屋に入って右に見える、首だけのロボットがギョロリだよ。
 08: 首がグネグネ動いて物を指したり出来るんだ。
 09: それから、ギョロリの反対側にはプロットという装置があるよ。
 10: キャラクターを空間に投影して自由に動かすことが出来るんだ。
 11: そして、目の前のほくの名前は、プログロロボット。
 12: 様々なプロクをジェスチャを交えて読めるよ。
 13: 学生が使う部屋はもう一部屋あって、面白いロボットがあるよ。
 14: その部屋についてはまた今度話すね。ハイバイ。

図 6: Long 型文

01: みなさん、こんにちは
 02: 今から安西・今井研究室を紹介するよ。
 03: この部屋は、学生たちが主にデスクワークする場所なんだ。
 04: だから、あちこちがパソコンだらけになってるよ。
 05: パソコン以外には、ユニークなロボットが沢山あるよ。
 06: いくつか紹介するね。
 07: まず、右に見えるのがギョロリだよ。
 08: 首が動いて物を指したり出来るんだ。
 09: それから、左手にはプロットがあるよ。
 10: キャラクターを投影して自由に動かすことが出来るんだ。
 11: そして、ほくの名前は、プログロロボット。
 12: プロクをジェスチャを交えて読めるよ。
 13: 別の部屋にも、面白いロボットがあるよ。
 14: その話はまた今度話すね。ハイバイ。

図 7: Normal 型文

4 動作イメージ

本章では、実際の動作イメージについて述べる。動作の流れは図9のとおりである。まず、ブログの記事ではメインで用いるGPS情報を設定する。これは、複数のGPS情報が同一記事内に存在した際に、ユーザが現場にいるかいないかの判定に使われる。記事中のメインGPS情報と読み手との距離が一定値以上なら、記事の文はLong型へと変換される。反対に距離が一定値未満なら、記事の文はNormal型へと変換される。さらにNormal型の状態においてPDAから追加命令があれば、記事の文はShort型へと変換される。例えば、BlogRobotが記事を再生する際、ユーザがボタン操作をすることでShort型へと切り替わり、すばやく情報を伝達することが可能である。判定部からユーザの状況に合わせて文章のモデルを変化させるシステムをFGS(Frequency of gesture and speech)と呼称する。

5 考察

本章では、3.3で挙げたスクリプトをもとにモデル変換に関する評価・考察を行う。

5.1 文章の再現性

3つのスクリプトを比較して、すぐ分かることとしてLong型はより詳細に、また具体的に書かれていると

01: こんにちは
 02: ここを紹介するよ。
 03: ここは、デスクワークする場所なんだ。
 04: だから、パソコンだらけだよ。
 05: それ以外には、ロボットが沢山だよ。
 06: 紹介するね。
 07: まず、あれがギョロリだよ。
 08: 首が動くよ。
 09: それからあっちにはプロットがあるよ。
 10: キャラクターを投影することができるんだ。
 11: そして、ほくはブログロボット。
 12: ブログを読むよ。
 13: 向こうの部屋にも、ロボットがあるよ。
 14: それはまた今度ね。バイバイ。

図 8: Short 型文

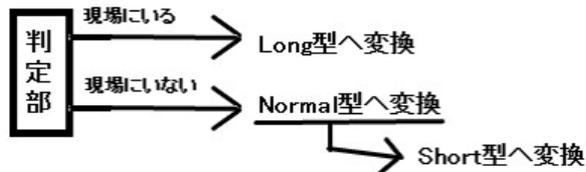


図 9: FGS の動作イメージ

いう点があげられる。例えば、L の 2 つめの文において単に「研究室」だけでなく、「慶應義塾大学」という修飾が追加されている。また、N の 3 番目の文中に「この部屋」という表現があるが、L では「部屋番号 25-419」と特定可能な具体的な情報を述べている。この特徴は、N や S から L へ変換できないことを示している。仮に N や S から L に変換することを考えた時、存在しない情報を付け加える必要があり、変換は不可能である。反対に L から N や S に変換することは、元々ある情報以上に情報は増えないので現実的であると思われる。

しかし、Long 型の文章から N・S に変換するというのは、文章を必要な分だけ削除する必要があるということである。どのような箇所を削るかにに関して、十分に考慮する必要がある。

5.2 指示語の選択

文章の補完性から、Long 型は大変優れていると言える。しかし、N 型・S 型における最適な指示語の選択（例えば、「それ」「あれ」「これ」）は、人間の感性によるところが多いので L 型から変換するのは困難だと言える。

5.3 ジェスチャの再現性

S 型においてはジェスチャをメインに用いているので L 型における文章の再現性と同様のことが、ジェスチャに関して言える。S 型では、N 型から変換される過程で消えた文章をジェスチャで補うことを目的とし

ているが、削除される文章をどのようなジェスチャに変換するか、文章を理解しなければならず困難である。

6 まとめ

本稿では、BlogRobot が記事を再生する際、ユーザの現在位置によって記事中の発話文章やジェスチャ命令を変化させるシステム FGS に関する検討を行った。文章の再現性の観点から L → N 変換のほうが、N → L 変換よりも現実的であるといえる。しかし、文章やジェスチャをどのように削るか、最適な指示語の選択をどのように行うかなどの問題がある。

今回、具体的な変換手法に関しては触れなかった。今後、具体的な変換手法を模索し、書き手の負荷をどれほど軽減できるか、またどの程度効果的に変換するか詳しく調べる必要がある。さらに Short 型に関しても読み手にどの程度効果的か実験を通して検討していきたい。

参考文献

- [1] 株式会社トランスメディア:
<http://mapii.jp>
- [2] 頓智ドット株式会社:
<http://support.sekaicamera.com/>
- [3] KDDI 株式会社:
<http://kazasu.mobi/>
- [4] 藤本健太, 大隈俊宏, 桑山裕基, 野田誠人, 大澤博隆, 篠沢一彦, 今井倫太: LIP: プログコンテンツを提示するロボットの行動生成システム, *The 23rd Annual Conference of the Japanese Society for Artificial Intelligence*, 1H1, 1 (2009)
- [5] 大隈俊宏, 藤本健太, 桑山裕基, 野田誠人, 大澤博隆, 篠沢一彦, 今井倫太: プログからロボットの動作コンテンツを生成するシステム TENORI の提案, *The 23rd Annual Conference of the Japanese Society for Artificial Intelligence*, 1B2, 3 (2009)