音声入力を用いた仮想空間操作問題における指示表現と対話の分類

大谷　章　　黒澤　義明　　竹澤　寿幸
広島市立大学　情報科学部　　広島市立大学大学院　情報科学研究科
{otani,kurosawa,takezawa}@nlp.its.hiroshima-cu.ac.jp

1. はじめに

近年，一般家庭におけるパソコンとインターネットの普及により，多人数参加型ネットワークコミュニティも浸透してきた．このネットワークコミュニティでは，仮想空間上のオブジェクトを操作して遊んだり，ネットワーク越しで他のプレイヤーとの会話を楽しむことができる．「セカンドライフ」というソフトはその代表である．このようなソフトでは，一般的にコントローラやキーボードによりオブジェクトの操作や会話を行う．しかし最近では，家庭用ゲーム機でも，直接マイクに話しかけて，別のプレイヤーと会話を行うことが可能になっている．つまり音声でネットワークを介して他のプレイヤーと会話をしながら，コントローラ等でオブジェクトを操作することができるようになってきている．

本研究では，音声入力を用いて，仮想空間上のオブジェクトの操作と他者との会話を同時に行うシステムを開発している．

本システムの中で重要な点が，ユーザが違和感なくシステムを使用できるということである．会話を始める前に「これから会話を始めます」や，会話を終了するときに「会話を終了します」などの切り替えとなる言葉が必要なシステムでは，ユーザは違和感なく使用することができない．そのため，入力音声が仮想空間上のオブジェクトに対する指示（以降アニメーション）なのか，他者との会話（以降チャット）なのかを判別することが重要になる．例えばアニメーションにおいて「～して」や「～しろ」などの表現で，チャットでは「～するわ」などの様々な表現でユーザが発話すると考えられる．多人数参加型ネットワークコミュニティにおいて，仮想空間上のオブジェクトは動いたりすることはできるが，思ったり，考えたりすることができない．「思う」や「考える」のような動詞は人の心情，思考を意味し，さらに動詞の中には主語や目的語によって心情，思考等を意味するものも存在する．このため，これらの動詞を含む発話はチャットに分類されると考えられる．

本研究では，上記の分類を行うために，語尾，動詞，動詞項構造シソーラス[1]という動詞が用例，格情報により意味を割り振られている辞書に記述されているフレームを素性とし，Support Vector Machine（SVM）を用いて精度を求めた．その精度から動詞項構造シソーラスの有効性について検証する．

2. 動詞項構造シソーラス

動詞には使い方によってさまざま意味を持つ．「寄せる」という動詞を例とする．

①　「車を壁に寄せる」
②　「彼に信頼を寄せる」

上記の 2 つの文において，①の「寄せる」は人の動作を意味しており，②では人の思考を意味している．このように動詞には主語や目的語の取り方によって，その動詞の意味が変わる．

動詞項構造シソーラスは，動詞を用例，格情報に基づいて意味分類を行い，それらの情報が記述されている辞書である．先ほどの「寄せる」について動詞項構造シソーラスに記述されている内容を一つ示す．

729,寄せる,動作主,が,,,対象,を,,1,,,,,着点,に,,,,,,,
車を路肩に寄せる,状態変化あり,位置変化,
位置変化（物理）,着点への移動,,
([動作主]の働きかけで)[1]が[着点]にいる状態になる

上記の内容を上から順に説明する．1 行目には，左から ID，見出し語，他は格情報で，1 は項番号（変数）である．2 行目から 3 行目にかけて，例文，大分類 1，大分類 2，中分類，小分類 1，小分類 2 となる．4 行目はフレーム名である．このフレームが動詞の意味の最下層に位置していたため，本研究ではこのフレームに注目して，分類精度を求める．

3. 研究目的

音声入力と仮想空間操作を合わせた研究は，近年行われている．例えば，音声により，仮想空間上のオブジェクトを操作するシステムとして傀儡[2]がある．また，音声を用いて操作するのはオブジェクトだけでない．音声により，施設の情報案内を行う対話システムの研究もされている．このシステムの代表として，「たけまるくん」[3]がある．

傀儡では仮想空間内のオブジェクトが完全に言葉を理解し，仮想空間内における状況を把握した上で，対話的に

動作することが一つの目的である．ゆえに，オブジェクトの操作しかできず，実際の多人数参加型ネットワークコミュニティのように，他者との会話ができない，ということである．そこで本研究では，音声入力を用いて，仮想空間上のオブジェクトの操作と，他者との会話を同時に行うことができるシステムの開発を試みる．

今回着目するところは，入力音声がオブジェクトに対する指示表現（アニメーション）と他者との会話表現（チャット）を判別することである．「たけまるくん」では，使われる質疑応答において，大きく分けて，たけまるくん自身に関係のある質問と情報案内に関係のある質問に答えることができる．例えば，「身長はいくつですか」という質問はたけまるくん自身に関係のある質問と見なし，返事を返す．また，「トイレはどこですか」という質問は情報案内に関係のある質問と見なし，地図などを表示して案内する．一見，自動的に判別しているように見える．しかし，「たけまるくん」では，一問一答形式で質疑応答を行う．そのため，質問文またはその質問のキーワードとそれに対応する応答文のデータベースが存在する．このデータベースの作成は人の手によって行われている．ゆえに，「たけまるくん」では，入力音声をたけまるくん関係と案内関係の質問に自動的に分類しているといえない．そこで本研究では，入力音声をアニメーションかチャットかに自動的に分類するモデルの作成を行う．

4. 動詞項構造シソーラスの導入

2 章の①，②の二つの文を比較する．以下に二つの文で使われた「寄せる」が動詞項構造シソーラスに記述されているフレームを示す．

a. ([動作主]の働きかけで)[1]が[着点]にいる状態になる

b. ([動作主]の働きかけで)([1]が[2]を評価することで)[1]の頭の中で[2]に対する評価が定まった状態になる

上記のフレームはそれぞれ，a は①の「寄せる」，b は②の「寄せる」に対応している．より詳細な意味を述べるならば，a，b の小分類 1 について，a の小分類 1 には**着点への移動**，b の小分類 2 には**心理的立場**と記述されている．ゆえに，a では人間の動作を意味し，b では人間の心情を意味している．このため，a のフレームを持つときの動詞を含む発話をアニメーションかチャットかに分類することは難しい．しかし，b のフレームを持つときの動詞を含む発話はチャットに分類されると考えられる．このことから，動詞項構造シソーラスを用いたときのアニメーションとチャットの分類の精度について検証していく．

5. 実験

本実験では被験者から発話データを収集し，得られた発話データをもとにアニメーションとチャットの分類精度を求めた．以下の節では，被験者から発話データを集めるために使用したシステムについて，実験の方法，収集した発話データに対する評価方法について述べる

5.1. システムの概要

本実験で発話データの収集のために使用したシステムの画面を図 5.1 に示す．このシステムの挙動として，キーボード入力により，図 5.1 の中央に配置されているオブジェクトを操作することができる．例えば，矢印キー(→)を押すとオブジェクトが右に動くことになる．

5.2. 実験方法

本実験では，被験者が実験を行っている風景を録画して，書き起こすことによって発話データを集めた．音声入力を用いなかった理由は，音声認識システムによる誤認識を避けるためである．アニメーション，チャットの集め方に関しては以下の節で説明する．

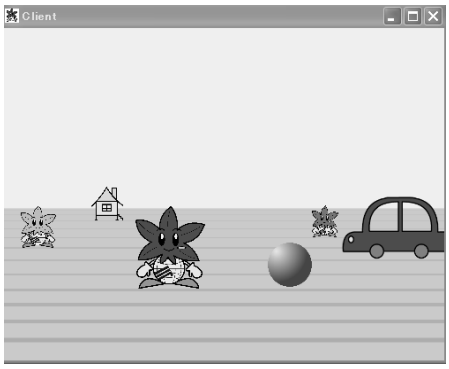

図 5.1：実験の画面

5.2.1. アニメーション

アニメーションの発話データを収集する方法として WOZ(Wizard of OZ)法 [4]を用いて，実験者がオブジェクトの操作を行うことにした．被験者には図 5.1 の中央に存在するオブジェクトに対して指示をしてもらい，その指示の内容にあった動作を実験者が操作する．例えば，「左に移動して」という発話に対しては，実験者がオブジェクトを左に動かすわけである． 被験者に仮想空間上で可能な動きについて口で説明したとき，そのとき発した言葉を被験者が使ってしまうかもしれない．このため実験開始前に，被験者には実験者が作成した仮想空間上でオブジェクトが可能な動きを行うビデオを見せた．

実験を行うとき，被験者にはあらかじめ課題をいくつか与えて，その課題を終了するような命令をしてもらった．例えば，「球を指定した場所まで配置してもらい，球の色を変えてもらう」という課題を画像，またはビデオを見せながら与えることになる．

5.2.2. チャット

チャットの発話データを収集する方法として, 2 種類の会話を用意した.

1 つ目は被験者と実験者が 5 分から 7 分間自由に会話する．2 つ目に，実験者と被験者は初対面という設定で，被験者と会話を行う. 例えば, 被験者は道がわからないので, 実験者に尋ねる, という背景のもと会話を行うことになる.

5.3. 評価方法

実験により 10 人分の発話データを収集することができた．この発話データに対して解析を行い，アニメーションとチャットの分類精度を求めた. 以下の節で精度を求めるまでの流れを説明する.

5.3.1. 発話データの解析

収集したすべての発話データに対して，アニメーションの発話にはアニメーションのタグを, チャットの発話にはチャットのタグを付けた.

次に形態素解析[5]，構文解析[6]を行った．なお本研究では，形態素解析の結果を一部変更した点があるので，そのことについて説明する.

「右に移動」などのサ変名詞が文末に現れたとき，形態素解析のときに自動的に「移動する」という動詞に解析するよう設定した．なお，活用形と文末情報に対して，情報はなしとした．文末情報の説明は次節で行う.

次に解析結果から得られた動詞すべてに対して, 動詞項構造シソーラスに記述されているフレームを割り当てる. 本研究では，この割り当ては手動で行うことにした.

5.3.2. 素性

文末情報(以降 sl)，動詞(以降 v)，動詞項構造シソーラスに記述されているフレーム(以降 t)を素性とした．この文末情報とは，「～して」や「～します」などの動詞の後ろに付く語尾のことである．しかし，「動け」という発話に対して，先ほど説明した文末情報はなく，「動く」という動詞の活用形には命令形という情報が存在する. このため, これら 3 つの素性にはそれぞれ活用形の情報も含まれている.

表 5.1：入力発話の解析結果例

例文	ボールの傍まで行って
タグ	アニメーション
v	行く
t	([動作主]の働きかけで)[1]が[着点]にいる状態になる
sl	て
活用形	連用タ接続

今回収集した発話の中で, 動詞の情報が含まれていない発話に関してはデータとして使わないことにした. 表 5.1 に発話データの例とその時の 3 つの素性と活用形の情報を示す.

5.3.3. 分類精度の求め方

本研究では，sl，v，t，slv(sl と v)，slt(sl と t)，all(sl と v と t)の 6 種類の素性の組に対して，SVM を用いてアニメーションとチャットの分類の精度を求めた. このとき SVM は TinySVM[7]を使用した．学習方法は発話データにアニメーションかチャットのタグ付けを行っているので，教師あり学習である．本研究では，まずクローズドテストを行った．この評価結果を表 5.2 に示す．表 5.2 の「シソ」は動詞項構造シソーラス，「文末」は文末情報，A は Accuracy, P は Precision, R は Recall である(A, P, R は以降も使用する)．表 5.2 から動詞項構造シソーラスを用いても，十分に学習されていないことがわかる.

表 5.2：クローズドテストの評価結果

	動詞	シソ	文末	slとv	slとt	slとvとt
	v	t	sl	slv	slt	all
A	92.8	89.6	86.1	97.4	95.2	97.9
P	90.5	86.9	83.6	97.4	94.0	97.8
R	91.8	87.2	81.6	96.2	94.0	97.1

6. 考察

表 6.1：交差検証による被験者の評価結果

	甲		乙	
	slv	slt	slv	slt
A	77.6	84.2	86.2	92.4
P	85.3	93.9	80.3	84.8
R	70.7	75.6	86.0	98.2

クローズドテストの結果だけでは, 動詞項構造シソーラスがどのような影響を及ぼしているのかがわからない. そこで, 各被験者の発話の特徴から動詞項構造シソーラスの効用について検討していく．そのために，交差検証法を用いて, 被験者ごとに分類を試みた. 表 6.1 に精度を記す.

この表は交差検証の際，テストデータとして扱った 10 人のうち 2 人甲，乙の slt，slv を素性としたときの結果である．動詞項構造シソーラスの採用により，精度の向上が著しい 2 名を選択した.

クローズドテストの結果（表 5.2）のように，一般的に slv に比べて slt の方が A，P，R が約 5%低くなる傾向が見られた．しかし，甲，乙では A，P，R すべてが上がった．以下では，動詞項構造シソーラスを用いて，精度が向上した理由について考える.

6.1. 着目する発話の条件

slv と slt において，システムによる分類が異なり，かつ slt でアニメーションに分類された発話について考える．条件に合った発話の中には「急ぐ」,「違う」を含む発話が複数確認された．このときのフレームはすべて「急ぐ」は F1,「違う」は F2 となる．F1 と F2 は以下に示す．

F1　([動作主]の働きかけで)[1]が[経路]を移動して別の位置にいる状態になる

F2　([1=人]の)[2=動作]が結果的に失敗に終わる

このため,「急ぐ」と「違う」が含まれる発話について扱うことにした．

6.2.　「急ぐ」を含む発話について

本節では，動詞「急ぐ」，フレーム F1 を含み，6.1 節で示した条件に合う発話について考える．slv では，学習用データに「急ぐ」を含む発話は存在しなかった．そのため，語尾と活用形だけでチャットに分類された．しかし slt では，学習用データにフレーム F1 を含むアニメーションの発話が存在していた．そのため，アニメーションに分類されたと考えられる．

ゆえに，未学習の動詞に対して，動詞項構造シソーラスを有効に利用することができると考えられる．

6.3.　「違う」を含む発話について

本節では，動詞「違う」，フレーム F2 を含み，6.1 節で示した条件に合う発話について考える．学習データに存在する動詞「違う」はほとんどチャットで使われていた．このため，slv ではチャットに分類したと考えられる．学習データにおいて，チャットで使われた「違う」のフレームは以下のようになる．

F3　[1]が[2]と異なる状態である

また，アニメーションで使われた「違う」のフレームは F2 である．本節で着目する発話に含まれる「違う」のフレームも F2 である．F2 はアニメーションに含まれると学習することによって，slt ではアニメーションに分類されたと考えられる．

以上の観点から，動詞項構造シソーラスを使用すると，学習データにおける一つの動詞が片側に多く分類されても，その動詞のフレームから分類に影響を及ぼすと考えられる．

7.　おわりに

仮想空間上のオブジェクトを操作することのできるシステムを作成した．そのシステムを用いて，オブジェクトに対する指示表現（アニメーション）のデータと，被験者との対話データ（チャット）を収集した．収集したデータから，動詞の語尾，動詞，動詞項構造シソーラスに記述されているフレームを求めた．これら 3 つのデータを利用し，SVM を用いて，入力音声をアニメーションかチャットかに分類した．その上で，動詞項構造シソーラスを用いたときの有効性について検討した．その結果，動詞項構造シソーラスは有効に活用できるということがわかった．しかし，実験で使用した仮想空間操作のシステムでは，オブジェクトがある一定の動作しか行うことができない．そのため，どうしても同じ動詞が集まってしまうので，より様々な動作をオブジェクトが行える仮想空間を作っていかなければならない．

今後は，多くのデータを収集し，動詞項構造シソーラスをより効果的に活用していく．また，対話相手が人間かシステムかによって，音響的特徴に差が生じる[8]．例えば，発話時間などがある．このような特徴を用いることも重要であると考えられる．

参考文献

[1] 竹内孔一，乾健太郎，竹内奈央，藤田篤，“意味を包含関係に基づく動詞項構造の細分類，”言語処理学会年第 14 回年次大会，pp.1037-1040，2008.

[2] 新山祐介，徳永健伸，田中穂積，“自然言語を理解するソフトウェアロボット：傀儡，”情報処理学会論文誌，vol.42，no.6，pp.1359-1367，2001.

[3] 西村竜一，西原洋平，鶴身玲典，李晃伸，猿渡洋，鹿野清宏，“音声対話エージェントによる生駒市コミュニティセンターの案内システム，”情報処理学会第 65 回全国大会，2F-5，2003.

[4] Norman M. Fraser , G. Nigel Gilbert , “Simulating speech systems,” Academic Press Limited，vol.5，pp.81-99，1991.

[5] 松本裕治，“形態素解析システム「茶筌」，”情報処理，vol.41，no.11，pp.1208-1214，2000.

[6] 黒澤義明，市村匠，相沢輝昭，“シナリオを対象とした構文解析規則記述方法，”自然言語処理, vol.12, no.2，pp.25-62，2005.

[7] 工藤拓，山本薫，松本裕治，“Conditional Random Fields を用いた日本語形態素解析，”情報処理学会研究報告，NL-161，pp.89-96，2004.

[8] 伊藤敏彦，甲斐充彦，岩本善行，水谷 誠，由浅祐規，小西達裕，伊東幸宏，“目的地設定タスクにおける対話状況の違いによる言語・音響的特徴の比較，” 情報処理学会論文誌， vol.43， no.7，pp.2118-2129，2002.