## 条件付き相互行為と話し言葉の会話分析に関する一考察 ~場面別の対話システム向上を目指して~

A Consideration on Conditional Interaction and Conversation Analysis of Spoken Language - Toward Improving Dialog System by Scene -

○ 太田 博三1

Hiromitsu Ota 1<sup>1</sup>

<sup>1</sup>放送大学 教養学部 <sup>1</sup>The Open University of Japan

Abstract: In recent years, interactive systems and dialogue generation in natural language processing have attracted attention. Particularly due to the spread of chat bots to call centers, accurate human interactive response is required. On the other hand, qualitative interactions in sociology's ethnomethodology and discourse analysis / conversation analysis are beneficial. Therefore, once again, using the Japanese language learner conversation data corpus of the National Institute of Japanese Language, it is a consideration aiming to verify the effect and apply it to the tendency of dialogue collapse and dialogue generation.

## 1. はじめに

#### 1.1 研究の背景と目的

2020年のオリンピックに向けて、スマートスピーカーが家庭に普及し、自動運転が実用化されようとしている。人工知能が人間に代替し仕事を奪うとも言われている。既にPepper などのロボットやチャットボット (Chatbot) が商用化されている。Pepper はイオンモールなどでお店の案内をし、チャットボットはコールセンターの一部機能を代替しようとしている。ここで、データの数が足りないという切実な問題を有している。ディープラーニング(Deep Learning)による機械翻訳の精度は極めて高いが、対話応答の自動生成は不十分である。そこで本稿では、定性的なフレームワークのアプローチでサンプルサイズの不足分を補うことを試みたものである。

次に,エスノメソドロジーや会話分析での隣接対 などの定性的な規則や傾向を,対話応答に適用して みたいと考え,追及してみた.

現行の対話システムは IF 文による制御構造で開発されており、必ずしも隣接対や雑談の種類が明示的に、もしくは十分に活用しきれている訳ではない.

ディープラーニングの分野では、データの増幅 (Data Augumentation) が発達し、転移学習も少しずつ発展してきた. これらでスケール化させ、実用化に結びつける動きがあるものの、画像認識のコンピュータービジョンに留まっているのが現状である.

#### 1.2 研究の新規性

本研究の新規性はデータ量を増やすと大きな相違が生じるかを考察し、定性的な傾向規則性を対話システムに取り入れられるかを試みたことにある.

本稿では、誰もが入手可能なデータである日本語 学習者の会話データを用いることで、統計的な有意 性やサンプル数より、日常生活の感覚でわかること を重視したものである.次に、[質問]-[応答]や[申 し出]-[受諾/拒否]などの隣接ペアの類型が上記の データにどのくらいあてはまるかなど、計量化し、 実証的な知見が得られれば、対話応答の次のトピッ クや発話文の予測が可能となる.隣接ペア以外に、 主に以下の5つが有用と考えられる.

- 1) Yes/No の応答詞 : あー, うん, えー, そう
- 2) あいづち:んー,はい,はー,えー
- 3) 言いよどみ:んー,あー,えー

Copyright is held by the author(s).

The article has been published without reviewing.

- 4) 呼びかけ:ね,ねー
- 5) フィラー: あの一, その一, えーと, えっと 今後, これらを分析し, 構造化したものを, 対話 システム等に追加することで, ユーザーの満足度の 向上につながると思われる.

#### 1.3 研究の主な手法

基礎集計を中心に行いながら考察する.国立国語研究所の提供している「日本語学習者会話データベース」を用いて集計を行う.隣接ペアは本稿で定義する種類のものに限定し計量化する.次に、それらのペアが全体の会話の促進になっているかなどを考察する.また、その隣接ペアの前後、もしくは直後の発話が修復に向けてのものか、完全に破綻しているが強引に会話を続けたものであるのかも含めて、定性的な判断を行う.日本語学習者の能力評しているが強引に会話を続けたものであるの能力に変き上げ(Thrust up)」が用いられてランク上の「突き上げ(Thrust up)」が用いられているのが特徴である.この意図的に難しい面接試験での質問をすることで、外国人の日本語能力を見極めるという方法論を用いている.いわゆる、とっさの対応が本能とみなす類のものである.

#### 1.4 用いたデータセットについて

国立国語研究所が公開しているコーパスの中の1つである「日本語学習者会話データベース」(図1)を用いる.また KY コーパスも同様の趣旨で作られたものであり、適宜、用いた.1990年の入管法の改正により、日本の社会状況に応じて、外国人受入れの適切な方策が必要となり、日本語学習を必要とする住民(言語生活者)の需要に見合った言語教育の展開が期待されていた.ACTFL-OPI(全米外国語教育協会認定の面接式口頭能力テスト)を活用し、日本語を用いた自然な会話に限りなく近い対話で構成されている.

ACTFL(全米外国語協会)による OPI(Oral Proficiency Interview Test)に基づいており、日本語 OPI は 1993 年に発足し、15年近く経過している. ここでの判断尺度は、次の4つに区分されている.

- 1) 超級 (Superior)
- 2) 上級(Advanced)
- 3) 中級(Intermediate)
- 4) 初級 (Novice)

これは「日本語学習者会話データベースの利用 手引き(平成22年5月国立国語研究所)」によれば、言語運用能力は10種類の階級に区分されている(表1).対話のスクリプトは、インフォーマント(日本語学習者/データ提供者)とテスター(面 接者)とからなり、30分ほどの対話形式で構成されている。

また上記の10段階のOPIレベルや性別,年齢, 出身国などを選択することができる.検索条件を設 定してダウンロードすると,文字化(一部,音声 化)されたスクリプトが入手でき,有用である.

表 1-, OPI 能力区分表

区分	OPIレベル	階級	OPI評価
1	超級 (Superior)	1	超級
2	上級(Advanced)	2	上級-上
	<i>''</i>	3	上級-中
	//	4	上級-下
3	中級(Intermediate)	5	中級−上
	//	6	中級-中
	//	7	中級-下
4	初級(Novice)	8	初級−上
	//	9	初級−中
	<i>''</i>	10	初級-下

### 2. 先行研究

本考察では、下記の3つ区分した.1つ目は、エスノメソドロジーや会話分析などの社会学である.言語学も多分に含まれている.2つ目は、対話システムを支える自然言語処理、3つ目は、深層学習、すなわちディープラーニングである.

### 2.1 エスノメソドロジー・会話分析

坊農・高梨他(2009)では、隣接ペアとは、[質問] - [応答]の対をなす発話の連鎖を指すものとして、対話システムにおける対話モデルに発話連鎖構造の土台としているとある。さらに、隣接ペアの概念には、[質問]に対し、[応答]がなされなかった場合に、どのような修復連鎖や挿入連鎖構造が生起しながら会話が進行するかを述べている。魏(2015)は「あの一」や「まー」などをフィラーと定義し、発話者が何らかの心的操作を行っている最中に発するもので、場をつなぐ機能を持つ言葉と定義している。多くは「感動詞」や「間投詞」に区分される。このフィラーを使いこなすのも、あいづちなどと同じく、会話をつなぐ言葉として、留意したいと考えている。

#### 2.2 自然言語処理

対話システムに実装される可能性は示している. また, 徳永・乾・松本(2005)及び徳永(2014)は, チャット対話の収集からコーパス作成, そして チャット対話の構造モデルを提案している. この チャット対話の質問や返答などの談話機能を担う構 成単位が交換行為である. 交換単位は「働きか け」,「応答」,「補足」の3種類に区分され,さ らに2,3の枝葉に分かれている. また,素性に関す る考察は有益であり(表2),本研究ではこれらを 精緻化することが具体的な目標でもある.素性の組 合せと継続関係の同定や再現率は2人の場合でも3 人の場合でも、86%と高く、素性も厳選されてい る. 発言間の結束度は次の式で求めている. < n(名 詞), rel(助詞), v(動詞)>の共起確率 P(<n, rel, v>) を求める. この確率 P(<n, rel, v>)は, Probabilistic Latent Semantic Indexing(PLSI)で推定する. 単語の共 起を潜在的な意味から同時発生とみなす手法であ る. PLSI における共起確率 P(<n, rel, v>)は次の式で 与えられる.

 $P(<n, rel, v>) = \Sigma P(<rel, v|z) P<rel, v|z) P(n|z) P(z).$ ここで, z は共起の潜在的な意味クラス (隠れクラス)を指しており、パラメータの P(<rel, v|z>, p(n|z), p(z))は EM アルゴリズムで推定している.

表 2 素性一覧徳永・乾・松本(2005)

素性	素性の説明
発言の末尾の表層表現	各発話の末尾が句点, 読点, クエスチョンマークであるか否かの2値
CRRuとPREu間の発言時間の差	CRRuとPREu間の発言時間の差が2分以上であるか否かの2値
発言間の結束度	共起確率に基づくCRRuとPREu間とCRRuと NBNu s間の結束度の強い方を1とする2値
交換行為の対話クラス	対話行為辞典に各交換行為(20種類 ≒ 隣接対)のクラスに分類したもの
交換行為の末尾の表層表現	同一人物の複数読み取れる発話の一番最後 の末尾がクエスチョンマークであるか否かの2 値
交換行為の発言時間の差	CRRuとPREmの先頭の発話における発話時間の差が5分以上であるか否かの2値

#### 2.3 対話自動生成のディープラーニング

対話応答の自動生成に関しては、ICML Workshop(2015)で Vinyals et al(2015)の Google のチームが NIPS2014 で発表された Sequence to Sequence model を基としている.多層の Long-Short term memory (LSTM)を用いて文章をベクトル化(エンコード)し、別の多層 LSTM を用いてベクトルをデコード(復元)するものである.これは「日本語ー英語」間の機械翻訳でよく用いられているアーキテクトであり、従来と比べて、自然な会話を生成するようになった. Ghazvininejad et al(2018)は、上記のモデルを拡張発展させたものである.会話型だけでなく非会話型データも組み合わせることによ

り, Seq2seq における Neural Conversation Model を 発展させたものである.

## 3. 基礎集計と分析による考察

日本語学習者会話データベース全体的にデータを 見渡してみると、全データは 390 個ある. イン フォーマント(日本語学習者)の属性は、20 代が圧 倒的に多く、女性が男性の2倍近くおり、大半を占 めている. 日本語学校生や大学・大学院生が半分を 占めている(図1).

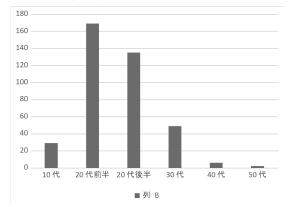


図3.1 インフォーマントの年代別分布

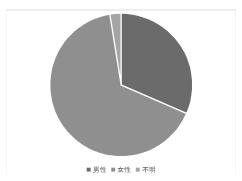


図3.2 インフォーマントの性別

#### 図3.3 インフォーマントの日本滞在時間

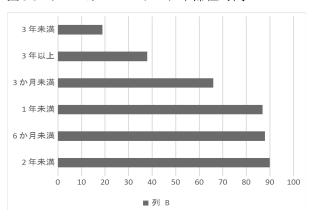


図3.4 インフォーマントの日本語学習期間

#### 3.1. 笑いとフィラーの比較検討

対話の中の笑いは、{笑}で表され、中級が笑(189)、超級は、笑(422)である.ここで、笑いの機能を検討してみると、早川(1997)や山根(2002)によると、会話のターンを維持・促進する働きがあるとされている.大きく以下の3つに分類される.

- 1) バランスを取るための笑い
- 2) 仲間づくりのための笑い
- 3) ごまかしのための笑い

一方,感動詞-フィラー(表 3.1.1)は、中級では、あー(189)に対して、超級では、あー(422)と半分以下になっている。フィラーは言いよどみや会話をつなぐ時間かせぎの機能もあり、超級の方が多様されており、会話をつなぐ言葉が多用されていると考えられる。

表 3.1.1 言いよどみ/ 戸惑いのフィラーの例

• • -		/ /		•	
辞書	文境界	書字形 (長層形)	語彙素	語彙素読み	品詞
現代語話1 言葉	I	ż	ネー	エー	威動詞-フィラー
現代語話1言葉	Ī	あのー	あの	アノ	威動詞-フィラー
現代語話1言葉	T	あ	あー	アー	威動詞-フィラー
現代語話1言葉	Ī	<b>あ</b> ー	あー	アー	威動詞-フィラー
現代語話1言葉	T	えー	えー	エー	威動詞-フィラー
現代語話1言葉	T	<b>ネー</b>	えーと	エート	威動詞-フィラー
現代語話1 言葉	T	あー	カー	アー	威動詞-フィラー
現代語話1言葉	T	<b>ネー</b>	えーと	エート	威動詞-フィラー
現代語話1 言葉	T	あの	あの	アノ	威動詞-フィラー
現代語話し言葉	T	あー	あー	アー	威動詞-フィラー
現代語話1言葉	T	あー	あー	アー	威動詞-フィラー
現代語話1言葉	Ī	あの	あの	アノ	威動詞-フィラー
現代語話1   宣葉	T	<b>ب</b> ــــ	٧	ŀ	威動詞-フィラー
現代語話1 言葉	T	<b>あの</b> ー	あの	アノ	威動詞-フィラー
現代語話し言葉	T	あのー	あの	アノ	威動詞-フィラー
現代語話1言葉	T	えー	えー	エー	威動詞-フィラー
現代語話1 言葉	T	<b>ネー</b>	えー	エー	威動詞-フィラー
現代語話1   言葉	T	h/	h-	ンー	威動詞-フィラー
現代語話1 言葉	T	<b>ネー</b>	えー	エー	威動詞-フィラー
現代語話1言葉	I	えー	えー	エー	威動詞-フィラー
現代語話1言葉	I	えー	<b>ネー</b>	エー	威動詞-フィラー
現代語話1 言葉	T	h/-	A/-	ンー	威動詞-フィラー
現代語話1言葉	I	その一	その	ソノ	威動詞-フィラー
現代語話1 言葉	T	h/-	4,-	·/-	威動詞-フィラー

#### 3.2 談話の構造分析の対話文生成への適用

隣接対をザトラウスキー(1993)は、応答ペアと呼び、1987年の国立国語研究所のをベースに、発話機能の定義と種類を12種類に分けている.①同意要求や②意思表示や③言い直し要求などである.筒井(2012)連鎖組織一覧として示しており、構造化されうるものと考えられる.また、発話機能連鎖の中には、山根(2002)のように、電話での談話などを、発話機能連鎖のパターン(表 3.5.1 及び表 3.5.2)とフィラーとが前後して生じていると考えることもできる.大浜(2006)は、日本語会話におけるターン交替形式を7種類に区分しており、定性的なモデルを提案している.

表 3.5.1 基本的な発話機能連鎖の中級の例

T: 社交機能(呼びかけ)	はい、はいじゃああの始めます、はじめまして
I: 社交機能(応答)	はじめまして一
T: 社交機能(自己提示)	ん私は【姓A】と申しますー
I:社交機能(応答)	あー【姓名B】申します
T:情報提供機能(呼びかけ)	あ【姓B】さんですね
I:情報提供機能(応答)	はい
T: 聞く機能+社交機能(呼びか	けんはい【姓B】さんはえーっと日本にいつごろ来たんです
I: 聞<機能+社交機能(応答)	ん、2か月ぐらい、前で
T: 社交機能(呼びかけ)	前、あーほんとう〈はい〉えーっとーん国はどちらですか
I: 社交機能(応答)	ん韓国人です
T: 社交機能(呼びかけ)	韓国〈はい〉あーそうですか韓国のどこーですか
I: 社交機能(応答)	釜山です
T: 社交機能(呼びかけ)	釜山〈はい〉あーそうですかー〈はい〉、んー日本に2か月
I: 社交機能(応答)	はい
T: 社交機能(呼びかけ)	んあそうですかんどうして日本に来たんですか
I: 社交機能(応答)	日本語勉強、する、つもりですけど〈んん〉、ん、ちょっと、
T: 聞く機能+社交機能(呼びか	けあ、そうですかー〈はい〉あーじゃあ文化も知りたい〈はい〉
I: 聞<機能+社交機能(応答)	建物とか私が関心、関心で〈ん〉、関心、持って、いた持っ
T: 聞く機能+社交機能(呼びか	<u>けあーそうですかー建物や車〈はい〉。はーそれは建物のデ</u>
<ul><li>I: 聞く機能+社交機能(応答)</li></ul>	はいけいじ〈けいじ〉、けいじゃいとか、大学で〈ん〉、建物
T: 聞く機能+社交機能(呼びか	<u>けあ造る〈はい〉、あ造る勉強をしていたんですか〈はいは</u>
I: 聞<機能+社交機能(応答)	あ建築はいはい{笑}
T: 聞く機能+社交機能(呼びか	けあーそうですかー〈はい〉え車もですか
I: 社交機能(応答)	{息を吸う音}車、も、好きですけど、まだ考え中です{笑}
T: 修復	んーあーそうですかー〈はい〉あじゃ建築の勉強してるー
I: 応答	はい

表 3.5.1 では、中級の例であるが、社交辞令での呼びかけと応答がしっかり成立している。一方で、「車」の話題で談話が破綻しているが、テスターが「建築」の話題に話を戻し、修復している。対話システムで、この修復の機能をいかに実装するできるかがを今後の課題としたい。

表 3.5.2 基本的な発話機能連鎖の超級の例

2(3.3.2.2	
T: 社交機能(呼びかけ)	こんにちは
I: 社交機能(応答)	こんにちは
T: 社交機能(自己提示)	はい私は【姓A】と申します〈あ〉お名前お願いできますか
I: 社交機能(応答)	あ【姓B】と申しますよろしくお願いします
T:情報提供機能(呼びかけ)	【姓B】さんですね〈はい〉はいどうも【姓B】さんはえあの一どちら のどちらからいらっしゃいましたか
I:情報提供機能(応答)	あ韓国から来ました
T:情報提供機能(呼びかけ)	あーそうですか〈えー〉, で, えー今日本では〈はい〉何をしてらっ しゃいますか
I:情報提供機能(応答)	今はえーとーパソコン関係の〈ん〉ウェブ関係のお仕事をしてま す
T:情報提供機能(呼びかけ)	あーそうですか〈はい〉ウェブ関係の仕事というとほんとに今の 〈えー〉現代の仕事という気がしますけれども〈はい〉じゃウェブを デザインしたりとかっということですか
I:情報提供機能(応答)	えーとーおもにあのモバイルサイト〈ん〉, のプログラムーを作ったりしてます
T: 同意要求(呼びかけ)	あーそうですか〈はい〉モバイル用の〈えー〉あのサイトということ ですね
I: 同意要求(応答)	はいはい
T:情報提供機能(呼びかけ)	あーそうですかでそれはあのとー会社ある会社のためにこうそう いったサイトを作るんですか
I:情報提供機能(応答)	そうですね
T:理解機能(呼びかけ)	{咳払い}ふーんそうですか
I:理解機能(応答)	はい
T: 同意要求(呼びかけ)	であの一今ねやはりなんでもかんでもがやっぱり〈えー〉あの一 モバイル時代〈えーえー〉そして携帯それからあとインターネット で全てまあ〈はい〉情報を得たりんとできるんですけれどもといろ いろな便利になってい〈分〈えーえー〉なんとなくコミュニケー ションのかたちが変わっていっているような気がするんですけど
I: 同意要求(応答)	そうですね~〈ん~〉かいしゃ~ [会社]社内でも, その一何かこう やっぱりプログラム~, してるときに〈え~〉大きい声で隅でしゃ べったりすると集中力が落ちるので〈ん~〉, みんな~メッセン ジャーで話したりとかするので
T:理解機能(呼びかけ)	同じ会社にいてもですか
I:理解機能(応答)	えーそうなんすよ〈わー〉, あのお隣の人でも〈えーえー〉メッセン ジャーでしゃべったりとかするので

表 3.5.2 では,まず,社交機能(呼びかけ)-社交機能(応答)があり,次に,情報提供機能(呼びかけ)-社交機能(応答)がなされ,理解機能(呼び

かけ)や同意要求(呼びかけ)とそれに対応して、理解機能(応答)や同意要求(応答)が成立しており、高度な対話がなされている.これらを、類型化したものが表 3.5.3 である.いずれも隣接対が多く見受けられ、またフィラーのようなつなぎ言葉も相手への同意を示すなど、発話促進の機能を果たしていると言える.

表 3.5.3 発話連鎖組織の一例

<連鎖5>	
OIA	情報要求= 語り要求
02B	情報提供= 語り(結論・評価)
03A	評価
04B	語り(詳細)
05A	理解・評価

## 4. データの増加の考察

本研究での新規性の一つととして,国立国語研究所の「日本語学習者会話データベース」(以下,国語研と略す)と同じ設計で取得された KY コーパスを追加することで,まずは、「品詞の分布」や「フィラー」,そして「笑い」の集計をし比較してみた.生産現場やコールセンターなどの商用の現場であると,定量的な根拠が求められる.少しでも,サンプルサイズについて勘どころが得られればと思われる.

#### 4.1 国語研のデータと KY コーパスの比較

表 4.1.1 に、国語研(339 個)と KY コーパス(44 個)のスクリプト数の一覧を示したものである. KY コーパスを追加すると、全体で 383 個となり、13%の増加となっている. 表 4.1.2 から、超級と上級の上のデータの追加が大きい. 出身国や年齢等の属性を加味しなければ、データの増加はマイナスにはならないと言える.

表 4.1.1

	国語研	KY-corpus	国語研+KY-corpus	増加率
超級	9	10	19	111%
上級-上	24	18	42	75%
上級-中	34	0	34	0%
上級-下	52	0	52	0%
中級-上	68	4	72	6%
中級-中	84	2	86	2%
中級-下	36	4	40	11%
初級-上	21	2	23	10%
初級-中	10	2	12	20%
初級-下	1	2	3	200%
	339	44	383	13%

図 4.1.1 国語研のスクリプト数の分布

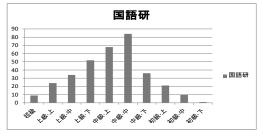


図 4.1.2 KY コーパスのスクリプト数の分布

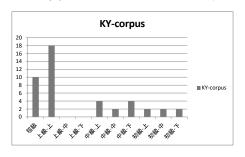


図 4.1.3 KY コーパスの追加差分(赤棒)

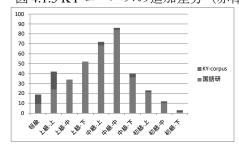
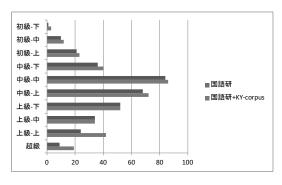


図 4.1.4 KY コーパスの増加分の比較



# 4.2 コーパスを増やす前と後との品詞の数量での比較考察

まず、表 4.2.1 に、国語研の上級者のスクリプトを形態素解析した後の品詞上位 30 個を示す.

表 4.2.1 国語研超級

1 補助記号-読点	6223
2 助動詞	6061
3 助詞-格助詞	5622
4 補助記号-括弧開	5350
5 補助記号-括弧閉	5348
6 名詞-普通名詞-一般	4822
7 感動詞-一般	4445
8 副詞	4035
9 感動詞-フィラー	3327
10 動詞-一般	3118
11 助詞-終助詞	2855
12 動詞-非自立可能	2618
13 補助記号-一般	2411
14 助詞-接続助詞	2235
15 助詞-係助詞	2109
16 助詞-副助詞	1691
17 記号-文字	1622
18 助詞-準体助詞	1412
19 名詞-普通名詞-サ変可能	1259
20 名詞-普通名詞-副詞可能	1218
21 代名詞	1159
22 連体詞	988
23 名詞-数詞	598
24 接尾辞-名詞的-一般	589
25 名詞-固有名詞-地名-国	553
26 形容詞-非自立可能	536
27 形容詞-一般	503
28 接頭辞	353
29 接続詞	327

30 名詞-普通名詞-助数詞可能 324

表 4.2.2

₹ <del>1</del> .2.2				
順位	基本形	個数		
1	ん	2150		
2	はい	2070		
3	です	1820		
4	か	1545		
5	の	1521		
6	とそう	1479		
7	そう	963		
8	スー	899		
	ŧ	894		
	んー	741		
11	Т	703		
	あの	677		
	た	671		
	な	634		
15	<u>し</u>	614		
16		614		
17	ます	566		
18	{	565		
19		565		
20	笑	506		
21	ない	487		
	その	464		
23		458		
	まあ	451		
25		449		
26	]	449		
27	_	428		
28	って	421		
29		407		
30	あ	384		

次に、KY コーパスの上級者のスクリプトを形態 素解析した後の品詞上位 30 個を示す.

表 4.2.3 KY コーパス超級 表 4.2.4

頁位 品詞	個数	順位 書字形(=表層形)	個数
1 助動詞	6364	1 です	174
2 助詞-格助詞	5959	2 うん	130
3 名詞-普通名詞-一般	4525	3 はい	916
4 感動詞-一般	4371	4 いう	914
5 補助記号-括弧閉	3378	5 そう	80
6 補助記号-括弧開	3377	6 あの	750
7 副詞	3221	7 ええ	70:
8 動詞-一般	3078	8 b—	56:
9 感動詞-フィラー	2707	9 あの一	56
10 動詞-非自立可能	2693	10 ます	53
11 助詞-終助詞	2658	11 ない	50
12 助詞-接続助詞	2407	12 えー	49
13 補助記号-一般	2268	13 から	44
14 助詞-係助詞	2029	14 その	42
15 記号-文字	1572	15 けど	37
16 助詞-副助詞	1571	16 それ	32
17 助詞-準体助詞	1371	17 まあ	31
18 代名詞	1331	18 h—	29
19 名詞-普通名詞-副詞可能	1277	19 こと	29
20 名詞-普通名詞-サ変可能	1221	20 こう	27
21 連体詞	1151	21 なん	25
22 接尾辞-名詞的-一般	647	22 笑い	24
23 名詞-固有名詞-地名-国	583	23 人	20
24 名詞-数詞	519	24 ちょっと	19
25 形容詞-一般	518	25 まし	19
26 形容詞-非自立可能	487	26 さん	19
27 名詞-普通名詞-助数詞可能	444	27 にほん	19
28 接続詞	405	28 ある	19
29 形状詞-一般	346	29 l.to	19
30 接頭辞	330	30 t=6	17-

最後に、国語研に KY コーパスを追加した後の上級者のスクリプトを形態素解析した後の品詞上位30 個を示す.

表 4.2.5 国語研+KY コーパス超級 表 4.2.6

-1	ПР	1 1 1 1 1 1 1 1
順位 品	詞	個数
1 助	動詞	12425
2 助	詞-格助詞	11581
3 名	詞-普通名詞-一般	9347
4 感	動詞-一般	8816
5 副	詞	7256
6 動	詞-一般	6196
7 感	動詞-フィラー	6034
8 助	詞-終助詞	5513
	詞-非自立可能	5311
	助記号-一般	4679
	詞-接続助詞	4642
12 助	詞-係助詞	4138
	詞-副助詞	3262
14 助	詞-準体助詞	2783
	詞-普通名詞-副詞可能	2495
16代		2490
17 名	詞-普通名詞-サ変可能	2480
18 連		2139
19 接	尾辞-名詞的-一般	1236
20 名	詞-固有名詞-地名-国	1136
	詞-数詞	1117
22 形	容詞-非自立可能	1023
	容詞-一般	1021
	詞-普通名詞-助数詞可	能 768
25 接		732
26 接	頭辞	683
27 形		625
	詞-普通名詞-形状詞可	能 542
29 形	状詞-助動詞語幹	340

順位 書字	₽形(=表層形)	個数
1 です	-	3564
2 はし	١	2986
3 いう		1902
4 そう		1764
5 あの	)	1427
6 えー	-	1392
7 うん		1327
8 ます	•	1096
9 んー	-	1040
10 あ <i>の</i>	)—	1019
11 ない	١	992
12 その		887
13 あー	-	877
14 つて	•	864
15 {		852
16 }		852
17 まあ	<b>5</b>	765
18 から	)	750
19 ええ		703
20 それ	ι	637
21 けと		611
22 こう		604
عے 23		589
24 人		528
25 笑	·	518
26 ちょ		486
27 なん	,	468
28 日本	Z	455
29 じゃ	·	391
30 45-	<b>パポル</b>	374

## 5. 突き上げ (Thrust up)と相互行為

## の考察

職場や医療現場,男女別の会話等の様々な場面に おける相互行為による研究がなされてきた.ここ で,OPI (Oral Proficiency Interview Test) での試験 監と外国人受験者での相互行為は,[名前の質問応 答]などの定量的なパターンマイニングによって, 以下のような、ある一定の談話構造がパターンが読み取れる.3つ目までは、あらかじめ、事前に対策して準備しておけば答えらえるものであるが、この間で試験管は、超級か上級か中級か、または初級かの4つに予測し分類し評価を行う.そして4つ目以降に、1つ以上難易度を高めた質問を行う.これが「突き上げ」である.受験者に対して、1対1の[質問-応答]や勧誘等を行い、即座の記憶ではない、内発的な日本語能力を評価する特殊や相互行為のであると考えられる.

また、ベイズ統計学のように、初めての受験者の場合と2度目、または3度目の受験者とでは、「突き上げ(Thrust up)」の効果は対数のように減少することが考えられる.いわゆる、条件付き相互行為と呼べるであろう.

1つ目: お互いの名前の[質問-応答],

2つ目: 出身国の[質問-応答], 3つ目: 趣味の[質問-応答]

4つ目:日本についてどう思うか?[突き上げ]

条件付き相互行為の例

例) P(2度目のインタビュー|1度目のインタ ビュー)

## 6. 今後の展望

本研究はチャットボットやコールセンターなどでの対話システムを対象としたものである。考察の中で興味深いのは次の2点であった。1つは、データの増加の影響とサンプルサイズについて、2つは、条件付き相互行為である。TOEICやTOEFLなどの同じ形式の対話では、1回目に比べて2回目が、2回目に比べて3回目が試験監の突き上げ(Thrustup)の効果が対数関数のような現象を示すということである。この効用関数はコールセンター等でも同じく確認できるものと考えられる。

## 参考文献

- [1] 国立国語研究所(2010)「日本語学習者会話データ ベースの利用手引き」
- [2] 鎌田・山内「タグ付き KY コーパス」 http://jhlee.sakura.ne.jp/kyc/corpus/
- [3] 徳永・乾・松本(2005)「チャット対話における発言 間の継続関係と応答関係の同定」 自然言語処理 言 語処理学会