

快適な情報享受を可能とする 音声対話システム

A Spoken Dialogue System for Enabling
Comfortable Information Acquisition and Consumption

2019年2月

高 津 弘 明

Hiroaki TAKATSU

快適な情報享受を可能とする 音声対話システム

A Spoken Dialogue System for Enabling
Comfortable Information Acquisition and Consumption

2019年2月

早稲田大学大学院 基幹理工学研究科
情報理工・情報通信専攻 知覚情報システム研究

高 津 弘 明

Hiroaki TAKATSU

目次

第 1 章	序論	1
1.1	研究背景	1
1.2	発話計画に基づく会話例	2
1.3	関連研究	3
1.4	本システムの特徴	5
1.5	システム要件と技術課題	7
1.5.1	負担の少ない表現への加工	7
1.5.2	興味・関心のある情報の提供	8
1.5.3	快適な会話リズムの実現	8
1.6	システム構成	8
第 2 章	主計画生成のための要約	10
2.1	序論	10
2.2	会話メディアとしての要約	11
2.2.1	話し言葉の同時性	11
2.2.2	報知性・指示性	12
2.3	要約コーパスの設計	12
2.3.1	Yahoo コーパス	12
2.3.2	日経コーパス	14
2.4	重要文抽出	16
2.4.1	単一文書要約と複数文書要約	16
2.4.2	重要文抽出モデル	16
2.4.3	単語の重要度	18
2.4.3.1	RandomForest を用いた単語重要度推定	18
2.4.3.2	BERT に基づく単語重要度推定モデル	19
2.4.3.3	補助情報	20
2.4.4	評価尺度 (ROUGE)	20
2.4.5	Yahoo コーパスを用いた評価実験	21
2.4.5.1	実験設定	21
2.4.5.2	実験結果	22
2.4.5.3	計算量	24
2.4.6	日経コーパスを用いた評価実験	25
2.4.6.1	BERT の事前学習	25
2.4.6.2	実験設定	25
2.4.6.3	実験結果	25
2.5	整列	27
2.5.1	関連研究	27

2.5.2	複数文書要約における整列アルゴリズム	27
2.5.3	Yahoo コーパスを用いた評価実験	28
2.6	文圧縮	30
2.6.1	関連研究	30
2.6.2	系列ラベリングモデル	31
2.6.2.1	Conditional Random Fields	31
2.6.2.2	Bidirectional LSTM	32
2.6.2.3	Pointer Networks	33
2.6.2.4	Differentiable Neural Computers	35
2.6.3	BERT に基づく文圧縮モデル	36
2.6.4	補助情報	36
2.6.5	Yahoo コーパスを用いた評価実験	37
2.6.5.1	実験設定	37
2.6.5.2	実験結果	38
2.6.6	日経コーパスを用いた評価実験	40
2.6.6.1	実験設定	40
2.6.6.2	実験結果	40
2.6.7	用語の説明に関して	42
2.7	要約の品質評価	43
2.8	まとめと今後の課題	46
第 3 章	会話によるニュース記事伝達のための口語化	47
3.1	序論	47
3.1.1	研究背景	47
3.1.2	関連研究	47
3.1.2.1	書き言葉から話し言葉への書き換え	47
3.1.2.2	「やさしい日本語」プロジェクト	48
3.1.2.3	対話システムにおける発話生成	48
3.1.3	想定するキャラクター	49
3.2	会話表現への書き換え	49
3.2.1	口語化の前処理:自然な文への復元	49
3.2.2	口調の書き換え	50
3.2.3	動詞・名詞・形容詞の書き換え	51
3.2.4	連体修飾節の展開	52
3.2.5	同格表現の書き換え	52
3.2.6	重文の発話分割	53
3.2.7	能動態 ⇒ 受動態, 他動詞 ⇒ 自動詞	54
3.2.8	名詞並列関係の書き換え	54
3.2.9	コメントの書き換え	55
3.2.10	その他の表現の書き換え	55
3.2.11	述語の時制推定	55
3.3	主観評価実験	56
3.3.1	実験設定	56
3.3.2	実験結果	58

3.3.3	違和感の理由	59
3.4	まとめと今後の課題	59
第4章	副計画生成	62
4.1	序論	62
4.2	副計画の生成方法	62
4.2.1	質問応答	62
4.2.1.1	固有表現に関する質問応答	62
4.2.1.2	why 型質問応答	63
4.2.1.3	定義型質問応答	64
4.2.1.4	真偽判定型質問応答	64
4.2.2	補足説明	66
4.2.2.1	類似文の提示	66
4.2.2.2	類似コメントの提示	66
4.2.2.3	情報の付加	66
4.2.3	その他	66
4.2.3.1	名詞反復確認	66
4.2.3.2	復帰点	67
4.2.3.3	同じような発話の回避	67
4.3	会話システムの応答分析	67
4.3.1	実験設定	67
4.3.2	実験結果	68
4.3.3	応答能力の実力値	72
4.4	まとめと今後の課題	73
第5章	ユーザー発話の意図理解	75
5.1	序論	75
5.2	関連研究	76
5.3	発話意図データセット	77
5.4	発話意図認識モデル	79
5.4.1	特徴抽出部の設計	79
5.4.2	識別部の設計	80
5.4.2.1	識別部 (P)	80
5.4.2.2	識別部 (L)	80
5.5	発話意図識別実験	81
5.5.1	特徴抽出部の学習と結果	81
5.5.2	識別部の実験設定	82
5.5.2.1	BERT の事前学習	82
5.5.2.2	補助情報	83
5.5.3	実験設定	83
5.5.4	実験結果	83
5.6	発話意図「質問」に関する予備実験	84
5.6.1	「質問」識別用データセット	84
5.6.2	LSTM の学習と識別結果	84
5.6.3	韻律情報と言語情報を用いた識別器の学習と識別結果	85

5.7	まとめと今後の課題	88
第6章	音声合成と間(ま)の制御	89
6.1	序論	89
6.2	関連研究	90
6.3	システムの構成	91
6.3.1	シナリオ生成部	92
6.3.2	発話テキスト解析部	92
6.3.3	音声合成部	92
6.3.4	ポーズ制御部	93
6.4	音声データの収録と分析	93
6.4.1	音声合成用音声データの収録と分析	93
6.4.1.1	発話単位	93
6.4.1.2	発話節の役割	94
6.4.1.3	収録原稿の作成	94
6.4.1.4	収録概要	95
6.4.1.5	発話節の位置と役割による韻律の違いに関する分析	96
6.4.2	ポーズ長推定用音声データの収録と分析	96
6.4.2.1	収録概要	96
6.4.2.2	データセットの分析	97
6.5	発話節の位置・役割を考慮した音声合成手法	99
6.5.1	発話テキストの解析	100
6.5.1.1	発話節への分割	100
6.5.1.2	発話節の位置・役割判定	100
6.5.1.3	アクセント付きかなテキストの生成	100
6.5.1.4	フルコンテキストラベルの生成	101
6.5.2	発話系列特徴を考慮した DNN 音声合成モデル	101
6.5.3	ポーズ長の推定	102
6.5.3.1	文節間ポーズ長推定モデル	103
6.5.3.2	文間ポーズ長推定モデル	103
6.6	評価実験 1: モデルの性能評価	104
6.6.1	音声合成	104
6.6.1.1	実験設定	104
6.6.1.2	実験結果	105
6.6.2	ポーズ長の推定	105
6.6.2.1	文節間のポーズ長の推定	105
6.6.2.2	発話文間のポーズ長の推定	108
6.6.2.3	ポーズ長の推定結果の例	108
6.7	評価実験 2: 情報を伝える話し方としてのふさわしさに関する主観評価	109
6.7.1	実験概要	109
6.7.2	実験結果	111
6.8	評価実験 3: 理解度に及ぼす影響の評価	112
6.8.1	実験概要	112
6.8.2	実験結果	113

6.9	評価実験 4: 質問しやすさに関する評価	114
6.9.1	実験概要	114
6.9.2	実験結果	116
6.10	質問を挟みやすくするための間（ま）の調整に関する予備実験	118
6.10.1	実験設定	118
6.10.2	実験結果	119
6.11	発話スタイルの転移学習に関する検討	121
6.11.1	実験設定	121
6.11.2	実験結果	122
6.12	まとめと今後の課題	123
第 7 章	統合システムの評価	124
7.1	情報伝達効率に関する評価	124
7.1.1	実験設定	124
7.1.2	実験 1: 自由対話タスクの実験結果	125
7.1.3	実験 2: 一問一答タスクの実験結果	129
7.1.4	自由対話と一問一答の情報伝達効率の比較	132
7.1.5	主計画のサイズを変えたときの 4 点以上の文の被覆率と 2 点以下の文の除外率の変化	132
7.1.6	実際の会話例	133
7.1.7	Naive な副計画との比較	135
7.1.8	質問応答の誤り分析	136
7.1.9	ユーザーの反応速度に関する比較	140
7.2	継続利用可能性と眠気防止効果に関する評価	141
7.2.1	実験設定	141
7.2.2	実験結果	141
7.3	主計画のサイズを変えたときの情報伝達効率に関する評価実験	143
7.3.1	実験設定	143
7.3.2	実験結果	143
7.4	まとめと今後の課題	147
第 8 章	結論	148
8.1	主計画生成のための要約	148
8.2	口語化	149
8.3	副計画生成	150
8.4	ユーザー発話の意図理解	150
8.5	音声合成と間（ま）の制御	150
8.6	統合システム	151
付 録 A	エラーハンドリング	152
A.1	序論・方法	152
A.2	主観評価	153

付 録 B トリビアの獲得	155
B.1 序論	155
B.2 関連研究	156
B.3 トリビアの識別手法	157
B.3.1 トリビア識別モデル	157
B.3.2 評価実験	158
B.3.2.1 実験設定	158
B.3.2.2 実験結果	159
B.3.3 Wikipedia からのトリビアの獲得	160
B.4 まとめと今後の課題	161
付 録 C オープンドメイン質問応答	162
C.1 序論	162
C.2 質問解析	163
C.2.1 固有表現	163
C.2.2 質問解析器の構築	164
C.2.2.1 素性	165
C.2.2.2 評価実験	165
C.3 ファクトイド型質問応答	166
C.3.1 回答候補の抽出	166
C.3.2 回答候補のランキング	167
C.3.3 意味的關係を用いた質問応答	169
C.4 why 型質問応答	171
C.4.1 原因部と結果部のチャンキング	173
C.4.1.1 因果関係候補文の抽出	173
C.4.1.2 形態素素性	173
C.4.1.3 構文素性	173
C.4.1.4 CRF によるチャンキング	174
C.4.2 回答候補の抽出	175
C.4.3 回答候補のランキング	175
C.5 定義型質問応答	177
C.5.1 CRF によるチャンキング	178
C.5.2 ランキング学習による定義文／説明文らしさの評価	179
C.5.3 「とは」を手がかり表現とした定義獲得	182
C.5.4 「類」の抽出	183
C.6 ニューラルネットワークを用いた回答段落選択	184
C.6.1 QA データの収集	187
C.6.2 実験設定	187
C.6.3 実験結果	188
C.7 まとめと今後の課題	188
付 録 D 要約コーパスの分析	189
D.1 日経コーパスの設計	189
D.2 日経コーパスの作業結果の分析	189
D.3 Yahoo コーパス vs. 日経コーパス	202

D.3.1	実験設定	202
D.3.2	実験結果	203
付 録 E	音声合成用音声データの分析	208
E.1	分析条件	208
E.2	分析結果	208
付 録 F	要約・平易化の従来研究	210
F.1	自動要約の従来研究	210
F.2	平易化の従来研究	217
付 録 G	言語資源・言語処理ツール	221
G.1	言語資源	221
G.1.1	ALAGIN の言語資源	221
G.1.1.1	動詞含意関係データベース	221
G.1.1.2	文脈類似語データベース	223
G.1.1.3	基本的意味関係の事例ベース	223
G.1.1.4	日本語異表記対データベース	224
G.1.1.5	上位語階層データ	224
G.1.1.6	日本語パターン言い換えデータベース	225
G.1.2	GSK の言語資源	226
G.1.2.1	京都大学格フレーム	226
G.1.2.2	拡張固有表現タグ付きコーパス	226
G.1.2.3	岩波国語辞典第五版タグ付きコーパス 2004	226
G.1.2.4	新聞記事 GDA コーパス 2004	226
G.1.3	その他の言語資源	227
G.1.3.1	EDR 電子化辞書	227
G.1.3.2	WordNet	227
G.1.3.3	日本語 WordNet 同義対データベース	227
G.1.3.4	現代日本語書き言葉均衡コーパス (BCCWJ)	227
G.1.3.5	日本語ウェブコーパス 2010	228
G.1.3.6	NAIST テキストコーパス	228
G.1.3.7	格助詞変換データ	228
G.1.3.8	「拡張固有表現 + Wikipedia」データ	229
G.1.3.9	Wikipedia 記事への促進・抑制関係付与コーパス	229
G.1.3.10	解答可能性付き読解データセット	229
G.1.3.11	Twitter 日本語評判分析データセット	229
G.1.3.12	評価対象-評価表現抽出用 日本語 Twitter データセット	229
G.1.3.13	単語感情極性対応表	230
G.1.3.14	日本語評価極性辞書	230
G.1.3.15	Polar Phrase Dictionary	230
G.1.3.16	日本語感情表現辞書	230
G.1.3.17	やさしい日本語コーパス	231
G.1.3.18	Simple PPDB: Japanese	231
G.1.3.19	日本語教育語彙表	231

G.1.3.20 基本語データベース 語義別親密度	231
G.2 言語処理ツール	232
G.2.1 形態素解析器	232
G.2.1.1 JUMAN / JUMAN++	232
G.2.1.2 茶筌 (ChaSen)	233
G.2.1.3 MeCab	234
G.2.2 係り受け解析器	234
G.2.2.1 KNP	234
G.2.2.2 CaboCha	235
G.2.2.3 J.DepP	236
G.2.3 述語項構造解析器	237
G.2.3.1 SynCha	237
G.2.3.2 ChaPAS	237
G.2.4 その他の言語処理ツール	237
G.2.4.1 Zunda	237
G.2.4.2 ASA	238
G.2.4.3 normalizeNumexp	239
G.2.4.4 意見（評価表現）抽出ツール	239
G.2.4.5 RaSC	239
謝辞	240
参考文献	241

表 目 次

1.1	本システムの特徴	6
1.2	システム要件と技術課題	7
2.1	選択文数と文圧縮による要約率	13
2.2	各作業者が選択した重要文タイプごとの平均文数	14
2.3	変数の定義	18
2.4	人間同士の比較結果の詳細	29
2.5	システムの正解率の詳細	29
2.6	異なる 2 者間で共通して選ばれた文の数	37
2.7	要約の各文に対する評価結果	44
2.8	要約の各文に対する評価結果 (カテゴリごと)	44
2.9	要約全体に対する評価結果	45
2.10	要約全体に対する評価結果 (カテゴリごと)	46
3.1	「情報のなわ張り理論」[神尾 90]における文末形式の分類	51
3.2	「現在」か「過去」か	56
3.3	「進行」かどうか	56
3.4	人間とシステムの総合平均点	58
5.1	発話意図の分類と対応するシステム動作	78
5.2	実験で使った発話意図データセットの統計	78
5.3	文脈として用いるシステム発話の範囲 (太字部分) (S がシステム発話, U がユーザー発話を表す.)	80
5.4	発話意図認識の実験結果	84
5.5	「質問」「非質問」データセットの統計	84
5.6	識別誤りの例	87
6.1	ポーズ長タグ付きコーパスの統計	97
6.2	収録音声から得られたポーズ長の例 (太線が発話文の区切り, 細線が発話節の区切り)	99
6.3	アクセント付きかなテキストの生成例	100
6.4	アクセント付きかなテキストの構成	101
6.5	発話系列特徴の有無による性能の比較	105
6.6	文節間ポーズ長推定モデルの結果 (RMSE)	106
6.7	ポーズ長の推定結果 (ミリ秒) (「推定結果の間」と「人の間」のいずれかが 100 ミリ秒以上であった箇所. 太線は発話文の区切り, 細線は発話節の区切り)	109
6.8	比較する合成音声の種類	110
6.9	【実験 3】タスク A : 間 (ま) の調整の有無に関する比較	113

6.10	【実験3】タスクB：単発話モデルと段落発話モデルの比較	113
6.11	【実験3】タスクC：発話系列特徴の有無に関する比較	113
6.12	質問しやすさの評価実験で用いたシナリオの例（太線は発話文の区切り，細線は発話節の区切り）	115
6.13	質問想定箇所を基準としたときの質問するまでの平均時間（秒）	117
6.14	発話スタイルを転移学習させるための補助情報の与え方	122
7.1	自由対話タスクの評価実験で使用したニュース記事のタイトルと文数	125
7.2	一問一答タスクの評価実験で使用したニュース記事のタイトルと文数	125
7.3	システム発話からユーザー発話までの平均時間（秒）	140
A.1	ローマ字表記の単語間の類似度計算の例	152
A.2	エラーハンドリングの評価で使用したシナリオ	153
B.1	トリビアデータセットの統計	159
B.2	トリビア／非トリビアの例	159
B.3	トリビア／非トリビアの識別精度（%）	159
C.1	本研究で扱う質問タイプ	164
C.2	RandomForest の設定条件	165
C.3	質問解析の性能	165
C.4	固有表現を用いた質問応答の例	168
C.5	獲得した意味的関係の例	169
C.6	意味的関係を用いた質問応答の例	171
C.7	原因／理由表現獲得のための正規表現ルール	173
C.8	構文ラベル	174
C.9	因果関係獲得のための IOB ラベル	174
C.10	why 型質問応答の例	176
C.11	定義文／説明文獲得のための IOB ラベル	179
C.12	定義型のスコア付け基準	180
C.13	説明型のスコア付け基準	180
C.14	定義型／説明型の質問応答の例	181
C.15	定義識別結果（RandomForest）	182
C.16	定義識別結果（線形カーネル SVM）	182
C.17	ラベルの種類	183
C.18	CRF による類の抽出結果（形態素素性）	184
C.19	CRF による類の抽出結果（文字素性）	184
C.20	CNF による類の抽出結果（形態素素性）	184
C.21	CNF による類の抽出結果（文字素性）	184
D.1	文圧縮作業結果：要約 A に関する各作業者の平均要約率（%）	196
D.2	文圧縮作業結果：要約 B に関する各作業者の平均要約率（%）	196
G.1	上位語階層データの例	224
G.2	Zunda の出力ラベル	238

目 次

1.1	主計画と副計画の例	3
1.2	想定する会話例	3
1.3	本システムの位置づけ	6
1.4	システム構成	9
2.1	会話利用を目的とした文圧縮の例 1	13
2.2	各作業者が重要文 a として選んだ文の位置の分布	15
2.3	各作業者が重要文 b として選んだ文の位置の分布	15
2.4	単語重要度推定モデル	19
2.5	訓練 1 人-テスト 4 人のときの結果	23
2.6	訓練 4 人-テスト 1 人のときの結果	23
2.7	MCKP+Rel の要約長ごとの入力文数に対する計算時間	24
2.8	MCKP 単体, Rel 単体での入力文数が 20 文固定のときの要約長に対する計算時間	24
2.9	重要文 a 抽出の実験結果	26
2.10	重要文 b 抽出の実験結果	26
2.11	整列の考え方	27
2.12	整列結果の正解率	29
2.13	単一記事で順序が入れ替えられたトピックの例	30
2.14	Bidirectional LSTM	32
2.15	オリジナルの Pointer Networks [Vinyals 15a]	33
2.16	Pointer Networks with Attention	34
2.17	Differentiable Neural Computers [Graves 16]	35
2.18	文圧縮モデル	36
2.19	訓練 1 人 - テスト 4 人 (全データ使用)	39
2.20	訓練 1 人 - テスト 4 人 (共通データのみ使用)	39
2.21	訓練 4 人 - テスト 1 人 (全データ使用)	40
2.22	文圧縮の実験結果	41
2.23	文法性を保証した際の文圧縮の実験結果	42
2.24	会話利用を目的とした文圧縮の例 2	43
2.25	要約の品質評価で用いたニュース記事の統計	44
2.26	要約の各文に対する評価結果	44
2.27	要約の各文に対する評価結果 (カテゴリごと)	45
2.28	要約全体に対する評価結果	45
2.29	要約全体に対する評価結果 (カテゴリごと)	46
3.1	口語化の前処理の例	49
3.2	アンケート用紙	57
3.3	各被験者の平均点のばらつき	58

4.1 ユーザー発話の分類と各発話タイプの割合	68
4.2 音声認識誤り発話に対するシステムの応答分類	69
4.3 「何それ？」に対するシステムの応答分類	70
4.4 「○○って何？」に対するシステムの応答分類	71
4.5 その他の質問に対するシステムの応答分類	72
4.6 現システムの応答能力の実力値	73
4.7 データベースを完備した場合の応答能力の理論値	73
5.1 発話意図ごとのラベラー間のラベル一致率 (Fleiss' Kappa)	78
5.2 発話意図認識モデル	79
5.3 CNN-AutoEncoder の出力例 (左:各スペクトログラムの5番目のスペクトルの周波数強度 (青:入力, 赤:出力), 中:入力したスペクトログラム, 右:出力したスペクトログラム)	82
5.4 発話意図「質問」の識別性能	87
6.1 システム構成図	91
6.2 発話節への分割例	93
6.3 収録で用いた原稿の例	94
6.4 役割ごとに見た発話節の段落内での出現位置の分布	95
6.5 発話文内の文節間のポーズ長 (非ゼロ) の分布	98
6.6 発話文間のポーズ長の分布	98
6.7 発話系列特徴を考慮した DNN 音声合成モデル	102
6.8 文節間ポーズ長推定モデル	103
6.9 文間ポーズ長推定モデル	104
6.10 文節間ポーズ長の誤差 (折れ線は累積相対度数)	107
6.11 文間ポーズ長の誤差 (折れ線は累積相対度数)	107
6.12 【実験2】タスク A: 間 (ま) の調整の有無に関する比較	111
6.13 【実験2】タスク B: 単発話モデルと段落発話モデルの比較	111
6.14 【実験2】タスク C: 発話系列特徴の有無に関する比較	112
6.15 【実験3】タイプ1の結果を1としたときの理解度の改善率	114
6.16 質問しやすさに関する評価結果	117
6.17 質問想定箇所を基準としたときの質問タイミングの分布	117
6.18 条件ごとに見た質問の事例数	118
6.19 質問に必要な間 (ま) の調査実験における質問の例	119
6.20 質問に要する時間 (被験者ごと)	120
6.21 質問に要する時間 (質問対象語ごと)	120
6.22 ポーズ長と質問カバー率の関係	121
6.23 発話スタイルの分類 [Iwata 11]	122
7.1 ユーザーが強く興味を持つ文 (5点をつけた文) を伝えられたかどうかに関する分析結果 (自由対話)	126
7.2 ユーザーが興味を持つ文 (4点以上をつけた文) を伝えられたかどうかに関する分析結果 (自由対話)	126
7.3 ユーザーが全く興味を持たない文 (1点をつけた文) を除外できたかどうかに関する分析結果 (自由対話)	127

7.4 ユーザーが興味を持たない文（2 点以下をつけた文）を除外できたかどうかに関する分析結果（自由対話）	127
7.5 ユーザー発話の種類と回数（自由対話）	128
7.6 ユーザーが強く興味を持つ文（5 点をつけた文）を伝えられたかどうかに関する分析結果（一問一答）	129
7.7 ユーザーが興味を持つ文（4 点以上をつけた文）を伝えられたかどうかに関する分析結果（一問一答）	130
7.8 ユーザーが全く興味を持たない文（1 点をつけた文）を除外できたかどうかに関する分析結果（一問一答）	130
7.9 ユーザーが興味を持たない文（2 点以下をつけた文）を除外できたかどうかに関する分析結果（一問一答）	131
7.10 ユーザー発話の種類と回数（一問一答）	131
7.11 自由対話と一問一答での情報伝達効率（EoIT; Efficiency of Information Transfer）	132
7.12 自由対話と一問一答における主計画のサイズ（ K ）を変えたときの 4 点以上被覆率と 2 点以下除外率の変化	133
7.13 実際の会話例（ S がシステム発話, U がユーザー発話, $\{S01, S02, S04, S05, S07, S09\}$ が主計画の内容）	134
7.14 Naive な副計画との比較（自由対話）	135
7.15 Naive な副計画との比較（一問一答）	136
7.16 被験者ごとの質問応答の種類と回数	136
7.17 被験者全体での質問応答の種類と回数	137
7.18 被験者ごとの質問応答の正解率（全体の平均正解率は 55%）	138
7.19 質問応答に関する音声認識誤りの割合（全体の平均音声認識誤り率は 20%）	139
7.20 継続利用可能性に関する実験結果	142
7.21 眠気防止効果に関する実験結果	142
7.22 情報伝達効率（全体）	144
7.23 情報伝達効率（要約 A）	145
7.24 情報伝達効率（要約 B）	145
7.25 主計画のサイズ（重要文 a の数 K ）を変化させたときの 4 点以上被覆率と 2 点以下除外率の変化	146
7.26 重要文タイプごとの抽出文数を変化させたときの情報伝達効率（合計抽出文数は 10 で固定）	146
A.1 エラーハンドリング評価実験の結果	154
B.1 トリビア識別モデル	157
B.2 テストセットにおけるトリビアと非トリビアの埋め込み表現 s の分布を多次元尺度構成法で次元圧縮し可視化した結果	160
C.1 質問応答システムの処理の流れ	163
C.2 構文情報のタグ付け例	174
C.3 回答段落選択モデル	185
C.4 Text Encoder	186
C.5 Embedding Encoder	186
C.6 Input Embeddings	187

C.7 実験結果	188
D.1 重要文抽出作業結果：1 トピックあたりの平均選択文数	191
D.2 重要文抽出作業結果：選択文数のヒストグラム（重要文 a）	192
D.3 重要文抽出作業結果：選択文数のヒストグラム（重要文 b）	192
D.4 重要文抽出作業結果：選択文数のヒストグラム（重要文 a+重要文 b）	193
D.5 重要文抽出作業結果：1 文目が選ばれたものの数	193
D.6 重要文抽出作業結果：選択文位置のヒストグラム（重要文 a）	194
D.7 重要文抽出作業結果：選択文位置のヒストグラム（重要文 b）	194
D.8 重要文抽出作業結果：1 文目を含めたものの頻度	195
D.9 重要文抽出作業結果：選択精度	195
D.10 重要文抽出作業結果：ROUGE-1	196
D.11 文圧縮作業結果：文節の一致率（要約 A）	197
D.12 文圧縮作業結果：文節の一致率（要約 B）	197
D.13 整列作業結果：記事の出現順と一致していたものの数	198
D.14 コーパスの記事の各段落を構成する文の数のヒストグラム	198
D.15 平均重要段落数	199
D.16 重要段落として選ばれた段落の数のヒストグラム（要約 A）	199
D.17 重要段落として選ばれた段落の数のヒストグラム（要約 B）	200
D.18 重要段落位置のヒストグラム（要約 A）	200
D.19 重要段落位置のヒストグラム（要約 B）	201
D.20 重要段落が包含する重要文の数のヒストグラム（要約 A）	201
D.21 重要段落が包含する重要文の数のヒストグラム（要約 B）	202
D.22 テスト：Yahoo コーパス，訓練：Yahoo コーパス	203
D.23 テスト：Yahoo コーパス，訓練：日経コーパス（要約 A）	204
D.24 Yahoo コーパスをテストセットとしたときの結果比較	204
D.25 テスト：日経コーパス（要約 A），訓練：日経コーパス（要約 A）	205
D.26 テスト：日経コーパス（要約 A），訓練：Yahoo コーパス	205
D.27 日経コーパス（要約 A）をテストセットとしたときの結果比較	206
D.28 テスト：日経コーパス（要約 B），訓練：日経コーパス（要約 B）	206
D.29 テスト：日経コーパス（要約 B），訓練：Yahoo コーパス	207
D.30 日経コーパス（要約 B）をテストセットとしたときの結果比較	207
E.1 発話節の位置による韻律の違い	209
E.2 発話節の役割による韻律の違い	209
G.1 JUMAN のカテゴリ情報	232
G.2 JUMAN のドメイン情報	232

概要

快適な情報享受は、質問応答のような意図性の高い能動的な情報行動だけでも、文書閲覧・聞き流しのような意図性のない受動的な情報行動だけでも実現できず、これらの情報行動のモードを自由に行き来することにより達成される。本研究では、意図性の変化を捉えるのに音声対話が適しているという立場から、ニュース記事を例題としてまとまりのある情報を利用者が快適に享受するための枠組みについて検討し、意図性の異なる多様な情報行動を高頻度に即応性高く切り替えながら情報にアクセスできる会話システムを開発した。

ここで提案するシステムは、主計画と副計画と呼ぶ、二種類の発話計画に沿って会話を進める。主計画は、ニュース記事など、伝達対象の文書を要約し、口語化することで作られるもので、文書の要点を伝えることを目的としている。ユーザーが受け身で聴いている限り、システムは主計画に従った情報伝達を行う。一方、副計画は、会話進行の各時点におけるユーザーの反応予測に基づいて、それらに対する応答を計画したもので、ユーザーの興味・理解度に応じた補足情報の提示を目的としている。ユーザーが情報要求を示してきた場合、システムは副計画に遷移し、補足説明を行う。これらの発話計画を事前に準備することで、目的とする即応性の高い円滑な会話インタラクションを実現した。システムの評価のために、情報伝達効率 (EoIT; Efficiency of Information Transfer) という指標を提案した。EoIT は、一連の会話によって、興味のある文を過不足なく効率的に伝達できたかどうかを図るための指標である。この指標を用いて評価を行った結果、要約のみの情報伝達や従来の一問一答型の対話システムに比べて、提案システムの方が情報伝達効率が顕著に高く、快適な情報享受が可能であることを確認した。

本論文の構成は以下の通りである。

第1章では、情報行動および音声の揮発性・同時性の観点から望まれる会話システムについて説明し、提案する会話システムの概要について述べる。

第2章では、会話によるニュース記事伝達のための要約手法について説明する。本システムでは、円滑な情報伝達を実現するために事前に発話計画を作成しておく。発話計画は主計画と副計画から構成される。主計画はニュースの要点に関する発話計画であり、ユーザーが受け身の姿勢で聴いている限り、システムはこの主計画に沿って発話する。主計画はニュース記事を要約した内容を基盤に生成される。ここで、要約は重要文抽出、整列、文圧縮の3段階の処理を経て行われる。重要文抽出は、文書から重要な文（核となる情報・見出し的な内容を含む文）

を数文抽出する問題である．重要文抽出の手法として最大被覆モデルを利用した．最大被覆モデルは，制限要約長に収まる範囲で重要な単語をできるだけ多く被覆するような文集合を抽出する手法である．そのため，単語の重要度の与え方が肝となる．従来，単語の重要度としては，頻度やTF-IDF，ロジスティック回帰で推定した値などが用いられてきた．本研究では，BERT (Bidirectional Encoder Representations from Transformers) を基本構造とした双方向モデルにより，文脈を考慮して単語の重要度を推定する手法を提案した．要約コーパスを作成し，評価実験を行った結果，従来の文脈を用いない手法よりも，提案モデルで推定した単語重要度を用いたときの方が高い ROUGE-1 を示した．整列は，重要文抽出で抽出した文の提示順序を決める問題である．要約の対象が単一記事である場合は，記事の出現順で問題ないことを確認した．複数記事からの要約である場合は，ある記事の重要文と別記事の重要文前後の文章との類似度に基づいて前後関係を決める手法を提案し，有効性を確認した．文圧縮は，文自体を短く要約する問題である．文圧縮の問題を文節レベルの系列ラベリングの問題として定式化し，BERT に基づく文圧縮モデルを提案した．実験の結果，単純な双方向 LSTM モデルに比べて，提案モデルの方が高い精度を示した．

第3章では，会話によるニュース記事伝達のための口語化手法について説明する．ニュース記事は書き言葉で書かれているため，会話で伝えるのに適した話し言葉に変換する必要がある．本研究では，友達と会話するようなカジュアルな感覚で利用できるシステムを目指している．そこで，敬体（ですます調）ではなく，くだけた表現に書き換える．中性的なキャラクターを想定し，ルールに基づく書き換えを行った．口調は「情報のなわ張り理論」に基づき伝聞口調に書き換える．その他，同格の展開，連体修飾節の展開，長い重文の分割などの書き換えを行う．口語化によって生成された発話文の言葉遣いに違和感がないか確認するため，主観評価実験を行った．実験の結果，システムが生成した発話系列に対する評価結果が，人間が口語化したものに対する評価結果と同程度であることを確認した．

第4章では，副計画の生成方法について説明する．要約で省かれた補足情報や予想される質問に対する回答を事前に副計画として用意しておくことで，ユーザーの情報要求に対して迅速な応答が可能となる．Wikipedia などを用いた用語の定義説明や，固有表現抽出結果を利用した日時，場所，人物などに関する回答計画，「ので」などの手がかり表現を利用した原因・理由に関する回答計画，含意・矛盾表現を用いた真偽判定，「何それ」といった抽象的な質問に対する回答計画を行う．質問応答として「何それ」に特化したシステムであることを被験者に伝え，応答の適切さについて評価したところ，68%程度の応答能力があることを確認した．副計画の事前設計による迅速な応答の有効性については7章の評価で確認した．

第5章では，システム発話の文脈を考慮した発話意図認識手法について説明する．より高い

情報伝達効率を実現するためには、ユーザーからのフィードバックを正しく理解することが重要となる。従来、発話意図の識別は、F0などの韻律情報に基づいて行われてきた。しかしながら、フィードバックに込められる意図をユーザー発話の情報のみから判断するのは難しい。そこで、本研究では、システム発話の文脈を考慮した発話意図認識モデルを提案した。本システムを用いて収集した対話データに対して発話意図のアノテーションを行ったデータセットを使用し、評価実験を行った結果、システム発話の文脈を考慮したモデルの方が考慮しないモデルに比べて高い識別精度を示した。

第6章では、談話構造上の発話の役割に応じたメリハリのある話し方を可能とする音声合成手法と間（ま）の推定手法について説明する。ニュースのようなまとまった量の情報を伝える会話システムでは、重要な情報がユーザーに伝わる話し方が求められる。この問題は、音声合成の分野では、これまで一文内におけるプロミネンスの付与という形で扱われてきた。しかしながら、発話の強調に影響を与える韻律は、談話構造に応じても変化することが古くから指摘されており、一文より長い単位で韻律や間（ま）を制御する必要がある。そこで、本研究では談話構造上の文の役割を、核（最も伝えたい内容を含む文）、前置き（核を伝えるためのリード文）、補足（情報を補足する文）の3つに分類し、これらをDNN音声合成の補助情報として利用することで、発話系列全体で韻律や間（ま）の制御が可能な音声合成システムを開発した。実験は次の3つの観点で行った。(1) 情報を伝える話し方としてのふさわしさに関する主観評価。(2) ニュースの理解度に関する客観評価。(3) 質問しやすさに関する評価。実験の結果、従来の1文単位の読み上げ音声で学習したモデルの合成音声よりも、情報伝達を意識させ段落単位で発話させた音声で学習したモデルの合成音声の方が、情報伝達のための発話としてふさわしいと感じるとともに、ニュースの理解度という面においても優れていることが分かった。さらに、発話の段落における位置や談話構造上の役割を補助情報に用いて合成することにより、メリハリのある聴きやすい話し方になることが分かった。また、発話の間（ま）に関しても、提案モデルで推定した間（ま）で発話した方が、短い単調な間（ま）で発話するよりも、聴きやすく頭に入りやすく、さらに質問しやすい話し方を実現できることが確認できた。

第7章では、提案システムと従来システムの比較実験結果について説明する。主計画と副計画からなるシナリオに基づいて、割り込みを許容しながら、質問に即座に回答できる点が本システムの特徴の一つである。このような即時応答機構を持つシナリオドリブンな提案システムと、要約を伝えた後にQAのモードに移行する従来の一問一答型の対話システムを比較評価した。その結果、情報伝達効率(EoIT; Efficiency of Information Transfer)という、提示できたほしい情報の被覆率と提示しなかったいらない情報の除外率の調和平均で定義される提案尺度において、提案システムの方が優れていることを確認した。また、本システムを継続して利用

したいかどうかについて、アンケート調査を行ったところ、過半数が継続して利用したいと回答した。

第8章では、全体のまとめと各要素技術の今後の課題について述べる。

付録A章では、音声認識誤りに対するエラーハンドリングについて説明する。音声認識誤りが原因で質問に適切に回答できないという問題が存在する。そこで、音声認識結果が記事の文脈から大きく外れるものであった場合、確認のプロセスを挟むというエラーハンドリングの機構を導入することで、質問応答の性能が向上することを確認した。また、エラーハンドリングありのシステムとエラーハンドリングなしのシステムで、どちらが良かったか主観評価を行ったところ、全ての被験者がエラーハンドリングありのシステムの方が良いと評価した。

付録B章では、トリビアの獲得方法について説明する。高い情報伝達効率を実現するにはユーザーからのフィードバックが必要不可欠である。フィードバックに寄与する要素の一つとして、ユーザーのコンテンツに対する興味がある。従来より、コンテンツに対する興味を誘引する戦略として、エンターテインメントや教育などでトリビアが利用されてきた。本システムでもトリビアの提示がコンテンツに対する興味を醸成する効果があると仮定する。文の構文に基づいてトリビアか否かを識別するモデルを提案し、97%と高い精度で識別できることを確認した。

その他、オープンドメイン質問応答について検討した内容を付録C章に記す。作成した要約コーパスの分析結果を付録D章に記す。収録した音声合成用音声データの分析結果を付録E章に記す。要約・平易化の従来研究に関する調査を付録F章に記す。本研究で使用した、または今後研究を進めるうえで役立つような言語資源・言語処理ツールについてまとめたものを付録G章に記す。

第1章 序論

ニュース記事のような文書で与えられたまとまりのある情報を，会話によって快適に享受する（受け取って自分のものにする，またその行為を楽しむ）ことを目的として，意図性の異なる多様な情報行動を高頻度に切り替えながら情報にアクセスできる会話システムを実現する．

1.1 研究背景

人間の情報に関する行動のうち，情報獲得・享受に関する行動は，意図的・能動的な情報行動と意図性のない受動的な情報行動に分類できるとされている [Case 07]．これまで，コンピュータサイエンスの領域では，主に前者の能動的な情報行動に焦点が当てられ，情報検索のような分野で活発に研究が進められてきた．一方，後者の受動的な情報行動は，偶発的な状況によることが多いため，研究要素に乏しいと考えられてきた．しかしながら，我々が日常で行う情報行動は両者の区別が明確ではなく，多様な意図性をもつ行動が複合したものであり，これらの状態を自由に遷移するものと言える．

特に，ここで対象とするニュース記事のようなまとまった量の情報を聴覚メディアで得ようとする場合，複合的な情報行動の扱いが必須となる．例えば，ニュースをラジオで聴く行為は，受動的な情報行動の典型例である．ラジオの聴取では，リスナーは受け身で聴いていればよいので楽に情報を取得できるという長所がある．しかし，途中で分からない単語が出てきても質問ができない，聞き逃した内容を聞き返せない，興味がない内容であっても最後まで聴き続けなければならないという短所がある．一方，対極にあるのが質問応答型の対話システムである．能動的な情報行動である質問応答は，断片的な情報を得るのには適しているが，まとまった量の情報を得ようとするとき，次々と質問を考えながら質問し続けなければならないため現実的でない．つまり，快適な情報享受は，質問応答のような意図性の高い能動的な情報行動だけでも，聞き流しのような意図性のない受動的な情報行動だけでも実現できず，これらの情報行動のモードを自由に行き来することにより達成されることが考えられる．

そこで，本研究では，受動的な情報行動を主体としつつも，情報補完のための能動的な情報行動を可能とする音声対話のモデルを実現し，これに基づいて快適な情報享受を実現する枠組みを提案する．また，単にこれらのモードを持つことにとどまらず，自由なタイミングで瞬時

にモードを遷移できる仕組みの実現を試みる。

ここで提案するシステムは、主計画と副計画と呼ぶ、二種類の発話計画に沿って会話を進める。主計画は、ニュース記事など、伝達対象の文書を要約し、口語化することで作られるもので、文書の要点を伝えることを目的としている。ユーザーが受け身で聴いている限り、システムは主計画に従った情報伝達を行う。一方、副計画は、会話進行の各時点におけるユーザーの反応予測に基づいて、それに対する応答を計画したもので、ユーザーの興味・理解度に応じた補足情報の提示を目的としている。ユーザーが能動的な情報要求を示してきた場合、システムは副計画に遷移し、情報の補完を行う。これらの発話計画を事前に準備することで、会話で重要な円滑なインタラクションを保証する。

まとまった量の情報を音声で伝える際に考慮すべきこととして、送受信の同時性と揮発性という音声対話の重要な特徴がある [畠 87]。文書による情報伝達では、書き手が全ての情報を伝え終えてから、読み手がその情報を受け取る。どのようにその文書を読んで理解するかは読み手の自由であり、文書のどの部分を、どんな速度で、何回繰り返して読むかは全て読み手に委ねられる。一方、会話による情報伝達では、話し手が情報を送信するのと同時に聞き手が情報を受信することになる。その瞬間ごとに聞き手が受信する情報は常に部分的かつ少量であり、聞き手はその情報の連続体を自分の力でまとめ上げ理解しなくてはならない。しかも、発せられた音声は瞬時に消えてしまう。そのため、まとまった量の情報を会話で伝えるのであれば、話し手は聞き手の負担を考え、一つ一つの発話に含める情報を極力抑え、分かりやすい表現で伝えるとともに、聞き手の疑問には即座に答え、これを解消しながら会話を進める手段を持つべきである。本研究では、副計画での補完を前提とした要約に基づく主計画の作成と、主計画文脈上でのユーザー応答予測に基づく副計画の作成という形で主計画と副計画を密に連携させ、動的な受動／能動のモード遷移を可能にすることで、このような会話に適した情報提示の仕組みの実現を試みる。

1.2 発話計画に基づく会話例

主計画と副計画の例を 1.1 に示す。また、この発話計画に基づく会話の例を 1.2 に示す。ユーザーが受け身で聴いている場合、システムは主計画の内容を伝えるので、 S_1, S_2 の発話を終えた後、 S_4 に遷移する。しかし、 S_2 の発話を終えた段階で、ユーザーが U_2 のような情報要求を示してきた場合、副計画に遷移し、 S_3 の内容を伝えてから、 S_4 に遷移する。これらのユーザー発話は、自由なタイミングで行うことができる。

ここで重要なことは、このような発話計画が事前に用意されているということである。もし仮に、ユーザーから質問がきた段階で、回答の探索や、会話表現への加工、音声合成などの処

理を行い、そこで多くの時間を費やしてしまうと、会話のリズムは崩れ、会話インタラクションによる情報享受の快適さは失われてしまう。本研究では、このような発話計画に基づくインタラクティブな情報伝達がリズムの良い会話を可能とし、それにより、ユーザーの情報行動が活性化され、ユーザーが欲する情報を効率的に伝達することが可能になるとの立場に立つ。

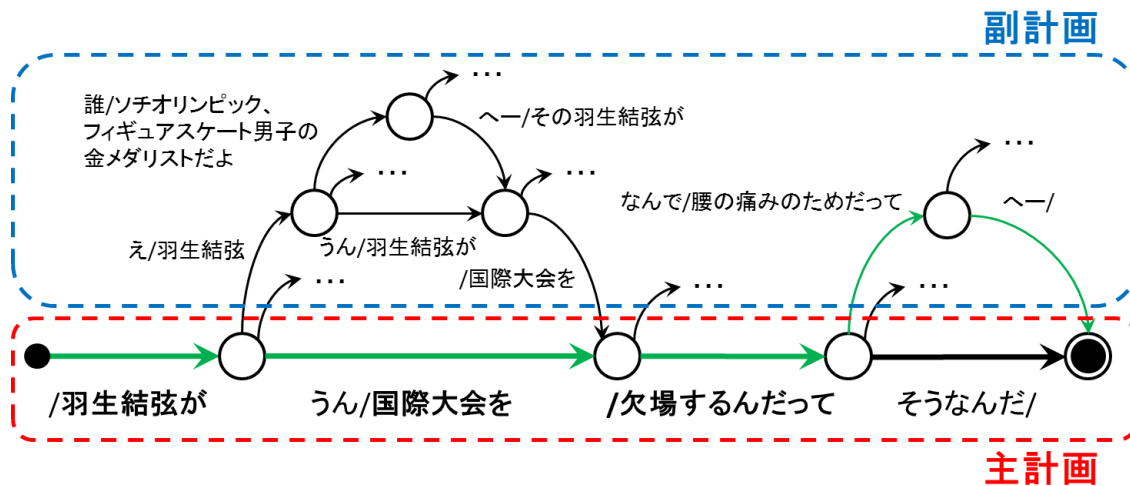


図 1.1: 主計画と副計画の例

S_1 :	羽生結弦が
U_1 :	うん
S_2 :	国際大会を欠場するんだって
U_2 :	え？なんで？
S_3 :	腰の痛みのためだって
U_3 :	へー
S_4 :	練習中に腰を痛めたそうだよ
	...

主計画(S_1, S_2, S_4) 副計画(S_3)

図 1.2: 想定する会話例

1.3 関連研究

音声対話は、本来は情報行動のモードを柔軟に遷移しながら進めるのに適したメディアであるが、現実にはSiri¹に代表されるように質問応答、すなわち能動的な情報行動に集中して研究・開発が進められてきた。先に述べた通り、このような質問応答型の会話は、断片的な情報を得

¹<http://www.apple.com/jp/ios/siri/>

るには適しているが、ニュース記事など、ある程度のまとまりをもつ情報を得る手段としては適さない。

Wikipedia やウェブニュースのようなまとまりのある情報を音声対話で伝える対話システムの研究がいくつかなされている。

水野らは、ユーザーにとって未知の情報を利用したシステム発話はユーザーを飽きさせないという仮説のもとで、日々更新されるウェブニュースを伝える対話システムを開発した [水野 09]。

Wilcock らは、Wikipedia の記事に関して、音声対話でトピックを検索し、記事内容を読み上げる WikiTalk と呼ばれるシステムを開発した [Wilcock 12, Wilcock 13, Wilcock 15]。このシステムには、ユーザーが次に何を話したいかを予測してスムーズにトピックをシフトさせる仕組みが組み込まれているが、トピックの内容に関する質疑応答といったトピック内でのインタラクションには対応していない。

杉山らは、ニュースのようなまとまった量の情報を音声で伝達する際に発生するであろう聞き取りや理解の失敗に着目し、ある程度の量を持つ情報を対話的に伝えることの有効性を示した [Sugiyama 00]。彼らは、特定の音声コマンド (e.g. 「もう一度」や「戻って」、「待って」) を使って、インクリメンタルに (小出しにして) 伝えられる情報をユーザーの理解に応じて操作できるような仕組みを取り入れ、単純に文章を読み上げた場合のシステムと比べて、必要な情報を獲得するタスクにおいてタスク達成率が高いことを示した。

翠らは、ユーザーとシステムの双方が主導権を取りながら、文書の情報にアクセスできる枠組みを提案し、観光案内のタスクで評価を行った [Misu 07]。彼らのシステムは、ユーザー主導の検索・質問応答モードと、システム主導の情報推薦モードを有しており、ユーザーの状態に応じて、これを切り替えながら、文書の情報を伝達する。まず、ユーザーから検索要求を受け付けた場合、該当する文書の情報を要約して伝える。次に、質問が来ればそれに答えるが、一定時間ユーザーの反応がなければ、システム主導の情報推薦モードに移行し、現在のトピックと関連した情報を提示する。特に、システム主導の情報推薦として、従来の文書の階層構造を利用した情報推薦や文書間類似度を利用した情報推薦に加え、ユーザーの興味を引き出すことを目的としたシステム主導の質問を提案しており、提案手法が他の手法に比べてユーザーによる受理率が高いことを示した。

吉野らは、ウェブ上のニュース記事を情報源として、雑談形式で情報案内を行う音声対話システムを開発した [吉野 14, Yoshino 15a, Yoshino 15b]。このシステムでは、まず、ニュースのヘッドラインをほぼ原文のまま読み上げる。ユーザーがそのトピックに興味を示した場合、ニュース記事の冒頭にあるサマリを読み上げる。もしここで、ユーザーが質問をしてきた場合

には、それに回答する。その後、ユーザーの反応が無くなっても、ユーザーがまだ興味を持っていそうな場合には、ユーザーとの対話履歴を用いてプロアクティブな情報推薦を行う [吉野 11]。

Traum らは、ホロコースト生存者の経験を質疑を交えながら伝える対話システムを開発した [Traum 15a, Traum 15b, Traum 16]。このシステムでは、ある事柄について一人称で語れる経験を伝える情報の対象としており、ニュースのような外的な情報を伝える本システムと扱っている情報の種類は異なるが、まとまりのある情報を伝えるという観点において共通している。

翠ら、吉野ら、Traum らの研究は、受動的な情報行動と能動的な情報行動を組み合わせる数少ない例ではあるが、これらにおいても情報行動のモードは明確に切り分けられており、情報行動のモード間を自由に遷移できるわけではない。そのため、まとまりを持つ情報を対象としてはいるものの、情報行動の自由なモード遷移という音声対話を持つ本来の快適さを十分に活用できているとは言えない。

1.4 本システムの特徴

快適な情報享受は「聞き手にとって負担の少ない形で、興味・関心のある情報を、即応性高く円滑に（リズムよく）伝える」ことで実現できると考える。このとき、ラジオや Siri といった従来の情報伝達システムと比較して本システムには表 1.1 のような特徴がある。本システムの利用姿勢は基本的には受け身であるが、ユーザーは能動的な情報要求を示すことでシステムから欲しい情報を取得できる。

対象とする情報は、ニュース記事のような情報の塊である。本システムでは、その情報の塊を咀嚼し構造化された発話計画に基づいて会話が進行する。情報の提供方法は、遭遇型と検索型のハイブリッドである。ユーザーは、遭遇型で新しい情報に触れつつ、検索型で気になる内容について詳しく知ることができる。

応答機構として、反射的応答機構と論理的応答機構を有する。反射的応答機構は、ユーザー反応の事前予測に基づいて用意された計画に基づいて即座に応答する機構である。一方、論理的応答機構は、計画外の質問に対して、その都度質問内容を吟味し、データベースから回答を探索し応答する機構である。特に反射的応答機構が本システムの特徴であり、これにより、即応性の高い迅速な応答を可能とし、リズムの良い会話を実現する。

また、本システムは音声コマンドによる端末操作や音声スケジューリングといった実用性を目指したシステムではなく、ラジオや音楽のような娯楽性を重視したシステムである (図 1.3)。しかしながら、その一方で、ラジオや音楽のように受け身に終始することなく、能動的な働きかけも楽しむことができるのが特徴である。さらに、本システムは、ハンズフリー、アイズフ

リー、ブレインフリー（思考を拘束しない）なシステムであるため、利用者は運転中や料理中など何か作業をしながらでも、作業を妨げることなく、気晴らしの手段として利用することができる。

表 1.1: 本システムの特徴

①聞き手にとって負担の少ない形で、②興味・関心のある情報を、③リズムよく伝える

観点	対応			本システム	参考	
	①	②	③		ラジオ	Siri
利用形態	○			受動利用＋能動利用	受動利用	能動利用
対象情報		○		情報の塊 → 知識（咀嚼して構造化）	情報の塊	情報断片
提供方法		○		情報遭遇＋情報検索	情報遭遇	情報検索
応答機構			○	論理的応答＋反射的応答	-	論理的応答

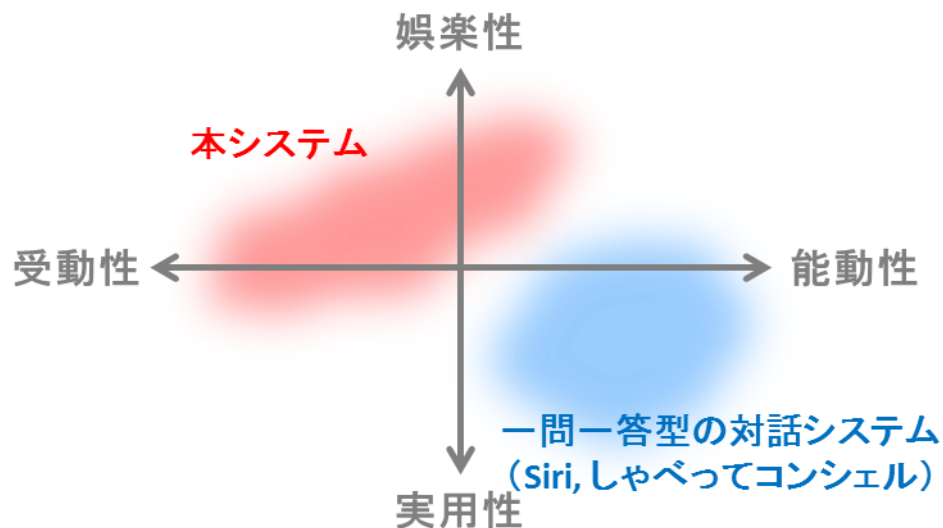


図 1.3: 本システムの位置づけ

表 1.2: システム要件と技術課題

技術課題	対応する章	システム要件		
		負担の少ない表現への加工	興味・関心のある情報の提供	快適な会話リズムの実現
主計画生成のための要約	第2章	✓	✓	
口語化	第3章	✓		
副計画生成	第4章		✓	✓
ユーザー発話の意図理解	第5章		✓	
音声合成と間（ま）の制御	第6章	✓		✓
エラーハンドリング	付録A章		✓	
オープンドメイン質問応答	付録C章		✓	

1.5 システム要件と技術課題

システムを次の3つの観点で設計する．各システム要件と技術課題の対応関係を表1.2に示す．

1.5.1 負担の少ない表現への加工

会話による情報伝達では、話し手が情報を送信すると同時に聞き手が受信することになる（話し言葉の同時性 [畠 87]）。そのため、その瞬間ごとに聞き手が受信する情報は常に部分的かつ少量であり、聞き手はその情報の連続体を自分の力でまとめ上げ理解しなくてはならない。しかも、発せられた音声は瞬時に消えてしまうため、文章のように何度も読み返すことができない。このように、会話による情報伝達では、話し手は聞き手に対してコミュニケーション上の大きな負担を強いることになる。一方、視覚メディアによる情報伝達を前提として書かれたニュース記事の一文一文には、たくさんの情報が埋め込まれている。もしその内容を会話で伝えるのであれば、話し手は送受信の同時性に起因する聞き手の負担を考え、極力情報量を抑え、分かりやすい表現で伝えなければならない。情報量を削減する手段として要約技術を用いる（第2章）。ニュース記事の文章は書き言葉で書かれており、堅苦しい表現が多く、会話で伝えるには適さない。書き言葉から話し言葉への変換として口語化処理を行い、会話に適した言語表現に書き換える（第3章）。会話において、生成された口語文は音声に変換して伝えるが、従来の音声合成器は、一文単位で合成されるため、発話系列全体として発話が単調で頭に入りにくいという問題がある。そこで、発話の談話構造上の役割を考慮したメリハリのある発話を可能とする音声合成システムを開発する（第6章）。

1.5.2 興味・関心のある情報の提供

ユーザーが受け身の姿勢で聞いている限り、要約された内容を伝えることになるが、会話の途中で興味や関心が変化したり、疑問が生まれたりすることは十分に想定される。こういった状態の変化が起きたとき、ユーザーは情報要求を示すことでシステムから欲しい情報を得ることができる。基本的には、事前に用意された副計画に基づいてユーザーの情報要求に応える（第4章）。しかしながら、全ての要求事項を事前に想定することは困難であるため、想定外の情報要求がきたときは、オープンドメインの質問応答機能でこれに応じる（付録C章）。ユーザーの情報要求が必ずしも明示的な言語表現として現れるとは限らない。場合によってはイントネーションなどの非言語情報も考慮する必要がある。そこで、韻律情報やシステム発話の文脈情報を用いた発話意図認識手法について検討する（第5章）。さらに、エラーハンドリングによる誤解解消の手続きを踏むことで、音声認識誤りに対して頑健な質問応答を実現する（付録A章）。

1.5.3 快適な会話リズムの実現

会話にはリズムがあり、適切なタイミングで発話することが求められる。期待されるタイミングで即座に発話するためには、その時点で既に発話の準備が完了していなくてはならない。そこで、本システムでは、会話で話す内容を事前に発話計画として用意しておくことで、円滑な情報伝達を可能とする。しかしながら、即座に発話が可能であったとしても、どのようなタイミングで発話すべきかという問題が存在する。早すぎるとユーザーの発話と衝突するリスクが増加し、遅すぎると情報享受の快適さが失われてしまう。そこで、質問しやすさおよび理解しやすさという観点で、間（ま）の調整を行う方法について検討する（第6章）。

1.6 システム構成

システムの構成を1.4に示す。本研究では、まとまりのある情報の例題としてニュース記事を扱う。計画生成部では、収集したニュース記事から発話計画を生成する。発話計画は主計画と副計画から構成される。主計画は、ニュースの要点に関する発話計画である。ニュース記事の内容を要約し、口語化することで生成される。副計画は、主計画の内容を補うための発話計画である。要約で省かれた情報や予測される質問に対する回答などを計画しておく。発話計画を事前に用意しておくことで、即応性の高い迅速な応答（反射的応答）が可能となる。生成された発話計画はデータベース上においておき、対話制御部でこれを参照しながら会話を進める。一方で、計画外の質問が来たときの対処方法としてオープンドメイン質問応答（論理的応答）機能を有する。ニュース記事の他に Wikipediaなどを知識源として質問応答用のデータベース

を構築する．音声認識には，Google Cloud Speech API² を使用している．

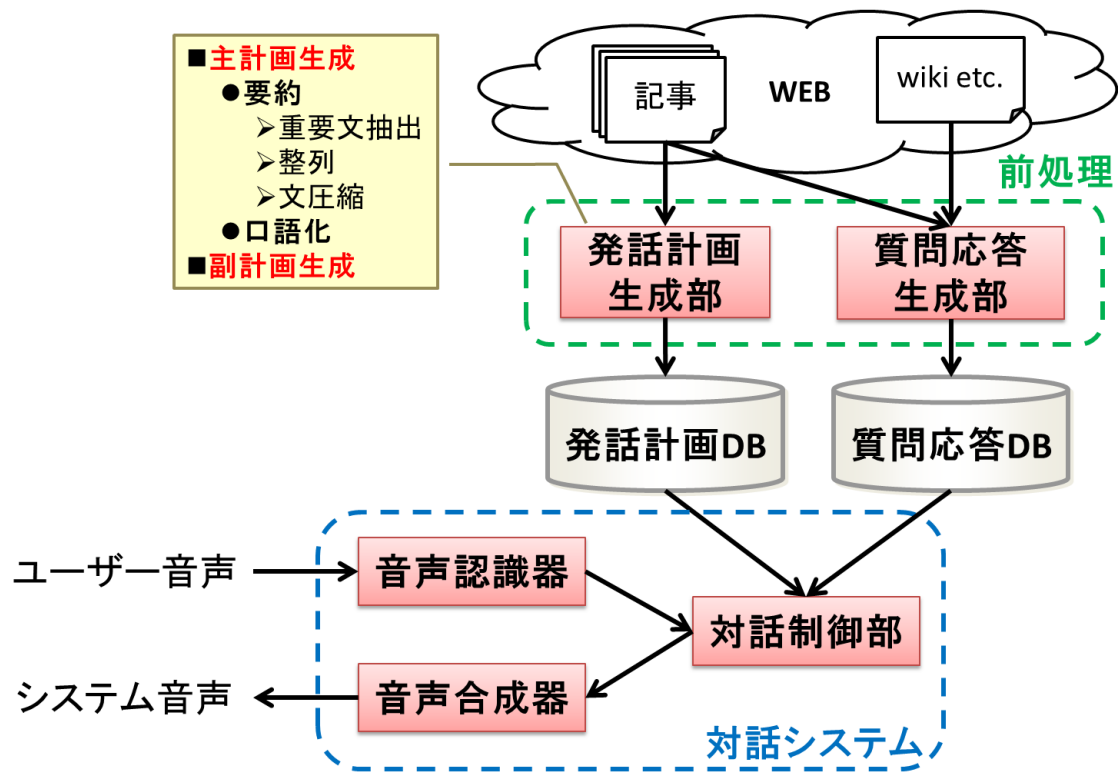


図 1.4: システム構成

²<https://cloud.google.com/speech/>

第2章 主計画生成のための要約

2.1 序論

BERT (Bidirectional Encoder Representations from Transformers) [Devlin 18] に基づく重要文抽出モデルと文圧縮モデルを提案する。

我々はニュース記事のようなまとまった量の情報を効率的に伝達する会話システムの開発を行っている。ここで「効率的」とは、伝達対象となる記事の中から、ユーザーにとって不要な情報を除き、必要な情報だけを伝えることを意味する。我々のシステムの特徴は、あらかじめ主計画、副計画と呼ぶ複数のシナリオを用意しておき、このシナリオに沿って会話を進めることで、リズムの良い会話を実現するうえで必須となる迅速な応答を可能としたところにある。主計画に沿って記事の要点となる情報を提示する傍らで随時ユーザーからのフィードバックを理解し、必要に応じて副計画に遷移して補足情報を提示する。このようにユーザーの興味や理解状態に応じて提示する情報を柔軟に切り替えながら会話を進めていく仕組みを持つ。

一方で、高い情報伝達効率を実現するには、ユーザーの興味や知識レベルに応じた主計画の生成が求められる。例えば、伝達対象のニュースに興味がありそうなユーザーに対しては大きめの主計画を、興味がなさそうなユーザーに対しては小さめの主計画を用意しておくことができれば、より効率的な情報伝達が期待できる。

主計画は、ニュース記事を要約することで作成する。一般的に、要約を作成する手法には、大きく抽出型 (Extractive) の手法と抽象型 (Abstractive) の手法がある。近年、抽象型の手法が盛んに研究されているが [Rush 15, Chopra 16, Nallapati 16]、文書の内容と無関係な単語を生成してしまう問題や文書の内容と矛盾する要約を生成してしまう問題がある。このような誤生成問題に取り組んだ研究はあるものの [See 17, Suzuki 17, 清野 18]、実用的なレベルには至っていない。これらの問題に加え、抽象型の要約モデルの学習には膨大なデータを必要とすること、我々のシステムでは要約で省かれた情報から副計画を作成するため、伝えた情報と伝えていない情報を区別する必要があることを加味し、本研究では抽出型の手法を採用する。

抽出型の手法として、近年では Encoder-Decoder モデルが用いられることが多い [Filippova 15, Cheng 16, Nallapati 17, Al-Sabahi 18]。これらのモデルでは、文や文書全体をエンコードして得られた埋め込み表現を用いて系列ラベリングの要領で文や単語を要約に含めるかどうかを

識別する．本研究では，ユーザーの興味に応じて情報量の異なる要約を作成することを目指しているため，重要文抽出においては，抽出する文の数をシステムのハイパーパラメータとして指定できる方が望ましい．そこで，本研究では，単語に重要度を割り当て整数計画法により重要な文を抽出する手法（最大被覆モデル）[Gillick 08, Takamura 09]を用いる．最大被覆モデルは，制限要約長に収まる範囲で重要な単語をできるだけ多く被覆するような文集合を抽出する手法である．そのため，単語の重要度の与え方が肝となる．従来，単語の重要度としては，頻度や TF-IDF [Filatova 04]，ロジスティック回帰 [Yih 07, Takamura 09] などで推定した値を用いることが多かった．

本研究では，BERT [Devlin 18] を基本構造とした双方向モデルで文脈を考慮して単語の重要度を推定する．これにより，従来の文脈を用いない手法よりも重要文抽出の性能が向上することを示す．また，文圧縮の問題は，文節単位の系列ラベリングの問題として定式化し，単純な双方向 LSTM を用いた場合よりも BERT に基づく提案モデルの方が優れていることを示す．さらに，同論文 [Devlin 18] で提案されている Masked Language Model と Next Sentence Prediction によりモデルの事前学習を行うことで性能が向上することを示す．

2.2 会話メディアとしての要約

自動要約の要約の従来研究については，付録 F.1 節にまとめた．

2.2.1 話し言葉の同時性

文章による情報伝達では，執筆者が文章を書き終えてから読者がそれを読む．つまり，情報の送信が終わってから，受信が行われることになる．そのため，読者は早く読もうがゆっくり読もうが，一部内容を見逃しように，途中で中断しようが自由である．一方，会話による情報伝達では，話し手が情報を送信すると同時に聞き手が受信することになる（話し言葉の同時性 [畠 87]）．そのため，その瞬間ごとに聞き手が受信する情報は常に部分的かつ少量であり，聞き手はその情報の連続体を自分の力でまとめ上げ理解しなくてはならない．しかも，発せられた音声は瞬時に消えてしまうため，文章のように何度も読み返すことができない．このように，会話による情報伝達では，話し手は聞き手に対してコミュニケーション上の大きな負担を強いることになる．

視覚メディアによる情報伝達を前提として書かれたニュース記事の一文一文には，たくさんの情報が埋め込まれている．もしその内容を会話で伝えるのであれば，話し手は送受信の同時性に起因する聞き手の負担を考え，極力情報量を抑え，分かりやすい表現で伝えなければなら

ない。そこで、情報を削減する手段として要約技術の利用が考えられる。

2.2.2 報知性・指示性

文章を簡潔にまとめる技術として要約技術がある。一般に要約は利用目的に応じて指示的要約と報知的要約に分類できる [奥村 04, 奥村 05]。指示的要約は、原文を読むかどうかの判断材料としての要約であるのに対し、報知的要約は、原文の代わりとして用いられる要約である。我々が開発している会話システムもユーザーが受け身の姿勢で聞いている限り、ユーザーは要約した内容のみを受け取ることになるが、ここで視覚メディアとしての要約と会話を志向した要約に大きな違いがある。会話で伝えられる要約はたとえ情報が不足していたとしても、送受信の同時性からユーザーはその場で情報要求を示すことで、瞬時に情報を補完できる。一方、報知的要約では、ユーザーが原文を見ることを想定していないため、提示される要約には情報の網羅性が求められる。また、指示的要約では、ある情報をもっと詳しく知りたいと思ったら、原文を参照して該当箇所を探しに行かなければならないのに対し、会話では聞き手からピンポイントで情報を取得することができる。このような送受信の同時性に伴う情報伝達の特徴から、会話では、一発話あたりの情報量は視覚メディアを対象とした要約よりも少なくなる傾向にある。

2.3 要約コーパスの設計

複数文書要約を含む Yahoo コーパスと、単一文書要約のみだが、2種類の重要文タイプからなる日経コーパスを設計した。前者は1記事あたり平均 10.64 文なのに対し、後者は1記事あたり平均 34.95 文である。

2.3.1 Yahoo コーパス

ニュースサイトから収集した 100 トピック（71 トピックが単一記事で、29 トピックが複数記事）について 5 人の作業者が要約を行った。Yahoo!ニュースを主な情報源として、この他、NHK ニュース、産経新聞、読売新聞、毎日新聞、nifty、Livedoor ニュースを複数記事要約用の情報源として利用した。また、トピックとして、オリンピックや選挙のような旬の期間が短いものは避け、豆知識となるような内容、比較的ポジティブな内容、新鮮さが失われにくい内容という観点から「IT・科学」を中心に選択した。作業者は、まず、重要文抽出を行い、ニュースの要点となる文を 2~4 文選択する。次に、抽出した重要文に対して整列を行い、会話として自然な流れになるように文の並び替えを行う。これらの作業に関して、落ちや教訓、今後の展望など会話の締めとなる情報が最後にくるような構成にするよう指示した。そして、それらの

文に対して、まず最初に会話で「こう伝えたい」という表現を作らせた後で、その口語表現を実現するために必要な情報を文節単位で選ばせた。文節の区切りはKNPで与えている。例えば、図2.1、図2.24のような原文が与えられたとき、会話で「こう伝えたい」という表現を先に考え（青色）、その口語表現を実現するために必要な情報を文節単位で抽出する（オレンジ色）。

コーパスの統計情報を表2.1に示す。選択された文数は1トピックあたり平均3文であった。文圧縮における文節数に基づく平均要約率はおよそ65%程度であった。「こう伝えたい」という口語表現と原文との文字数に基づく平均要約率はおよそ70%程度であった。後者の方が要約率が高いのは、口語表現では、やさしい表現を作り出すために同格表現や連体修飾節が「～っていう～」「～したんだけど、～」のように展開されることが多く、これにより選ばれた情報量に対して少し長めの表現が作られるためだと考えられる。

英国などの 国際研究チームは、 太陽系に 最も 近い 恒星 「プロキシマ・ケンタウリ」の 近くで、 地表に 水が あるかもしれな い 地球に 似た 惑星を 発見したと 発表した。
S : 地球に似た惑星が見つかったんだって
U : え？どこで？
S : プロキシマ・ケンタウリっていう恒星の近くだよ
U : 何それ？
S : プロキシマ・ケンタウリっていうのは、太陽系に最も近い恒星なんだ
U : へー

図 2.1: 会話利用を目的とした文圧縮の例 1

表 2.1: 選択文数と文圧縮による要約率

	選択文数	要約率（文節）	要約率（文字）
作業者 1	297	65.6	74.1
作業者 2	300	65.9	72.0
作業者 3	304	66.5	70.0
作業者 4	302	58.4	65.2
作業者 5	308	69.3	74.4

2.3.2 日経コーパス

日経新聞のテクノロジー系のニュース記事 100 個について 3 人の作業者に要約を作成させた。1 記事あたりの平均文数は 35 文である。要約作業は、重要文抽出、整列、文圧縮の 3 つの作業からなる。

重要文抽出は、ニュース記事の中から重要な文を数文抽出する作業である。ここで、重要文を 2 種類定義した。一つは、見出し的な内容・核となる情報を含む数文を抽出したものである。ここで抽出した文をタイプ a とし、タイプ a の文集合からなる要約を要約 A とする。もう一つは、要約 A の内容を補足するような文を数文抽出し、要約 A に加えたものである。ここで抽出した文をタイプ b とし、タイプ a とタイプ b の文集合からなる要約を要約 B とする。作業者には、タイプ a の文を 3 から 6 文、タイプ b の文を 2 から 5 文の範囲で抽出するように指示した。各作業者が選択した各タイプの重要文の平均文数を表 2.2 に示す。また、各作業者が重要文 a として選んだ文の位置の分布を図 2.2 に、重要文 b として選んだ文の位置の分布を図 2.3 に示す。その他の統計の詳細は付録 D 章に記す。

整列は、抽出した重要文を会話で伝えるのに適した順番に並び替える作業である。作業者には、落ちや教訓、今後の展望など会話の締めとなる情報が最後にくるような構成にするように指示した。

文圧縮は、抽出した重要文を会話で伝えるのに適した情報量に削減する作業である。ここでは、各文について会話でこう話したいという口語表現を最初に考えさせ、その口語表現を実現するのに必要な情報を文節単位で選択させた。

表 2.2: 各作業者が選択した重要文タイプごとの平均文数

	作業者 1	作業者 2	作業者 3
重要文 a	4.15	3.64	3.60
重要文 b	2.61	3.62	3.66

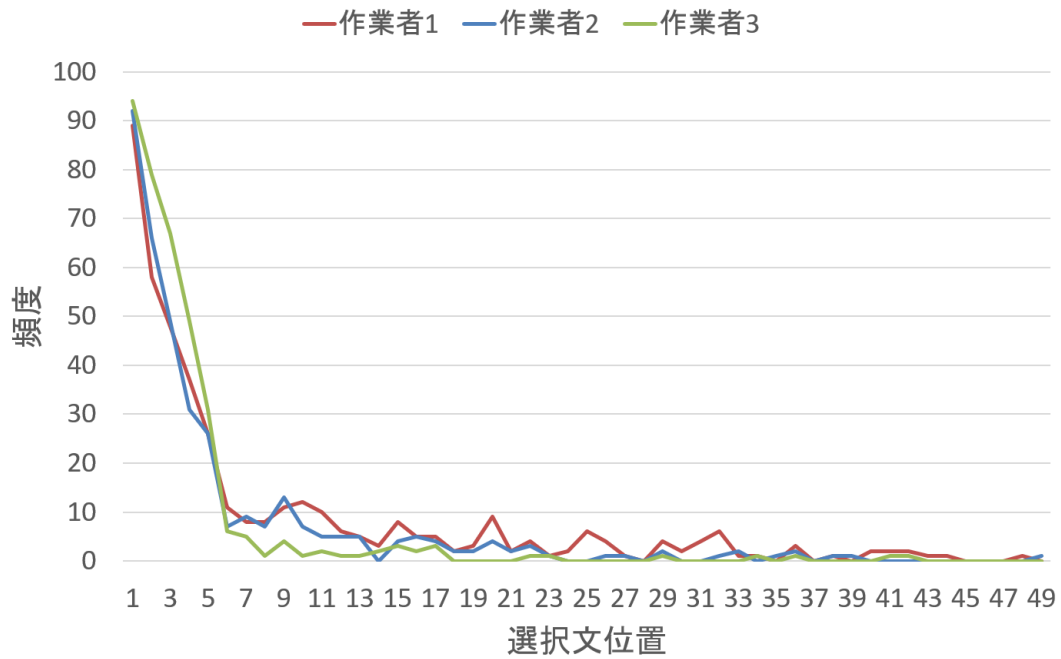


図 2.2: 各作業者が重要文 a として選んだ文の位置の分布

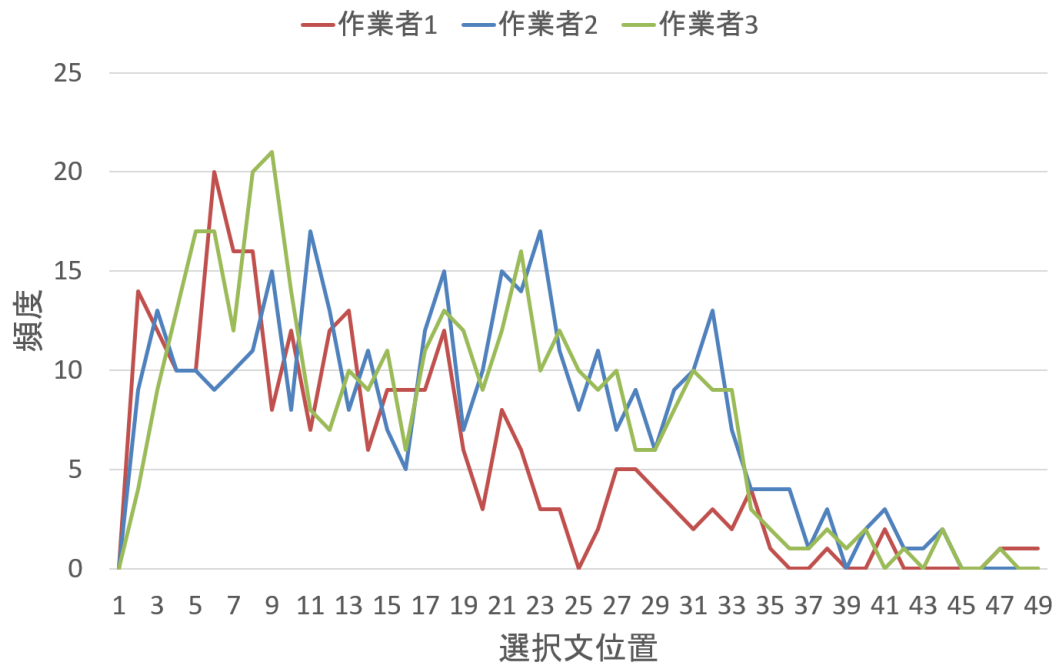


図 2.3: 各作業者が重要文 b として選んだ文の位置の分布

2.4 重要文抽出

文書閲覧のような視覚メディアによる情報アクセスでは、読み手は必要な情報だけを効率的に選びながら読み進めていくことができる。一方、聴覚メディアにはこのような拾い読みに相当する手段がない。そのため、ニュース記事の内容を未加工のまま伝えたとすると、聞き手はたとえ興味がない内容であっても最後まで黙って聴くしかない。そこで、ニュース記事の内容を要約して伝えることが考えられる。本節では、重要文抽出技術により、会話で伝えるべき重要な情報を文単位で抽出することを目的とする。

2.4.1 単一文書要約と複数文書要約

要約の問題は、入力とする文書の数によって、単一文書要約と複数文書要約に分けられる。単一文書要約の場合、入力全体で同じ内容が何度も出現することは少ないので、冗長な要約になる心配は少ない。一方、複数文書要約の場合、類似する内容が複数の文書に出現する可能性があるため、冗長性の制御が重要な課題となる。本研究では、多様な情報源を活用できた方が発話計画が充実することから、複数文書要約にも対応させる。

2.4.2 重要文抽出モデル

重要文抽出は、与えられた文書の中から要約に含めるべき内容を文単位で抽出するタスクである。一般に、複数文書要約の問題はナップサック制約付最大被覆問題 (maximum coverage problem with knapsack constraint; MCKP) として定式化されることが多い [Gillick 08, Takamura 09]。最大被覆問題に基づく要約モデル (最大被覆モデル) は、被覆した概念 (ユニグラムやバイグラム) の重要度の和が最大となる文集合を選択するモデルである。このモデルは、制限長内でできるだけ多くの内容を含めようとするため、自然と冗長性が削減される。しかしながら、文間の関係性や主題への関連性について考慮していないため、生成される要約は話の流れに一貫性を欠いたものになる恐れがある。一方で、McDonald は文書要約問題を主題への関連性から文同士の類似度を差し引いた目的関数で定式化している [McDonald 07]。このモデルでは、文同士の類似度を差し引くことで目的関数で直接的に冗長性を扱っているが、基本的には主題への関連度が高い文が好まれて選択されるため、複数文書に対して適用した場合、似たような文が選ばれる可能性が高い。

そこで、冗長性と内容の一貫性の問題を同時に扱うモデルとして、最大被覆モデル (MCKP) と McDonald のモデル (以降、関連性モデル; Rel) を組み合わせたモデル (MCKP+Rel) を提案する (式 2.2)。各変数の説明を表 2.3 に示す。要約の長さ K に関しては、目的に応じて文数

であったり、文字数であったり、バイト数であったりするが、ここでは、文数を用いる。式 2.4 が、 K 文以下の要約を作るための要約長に関する制約である。要約長を文字数やバイト数で規定する場合、変数 c は各文の文字数やバイト数となるが、ここでは、要約長を文数で規定しているため、文の長さによらず 1 とする。式 2.5 は単語 l を含む文が一つでも選択されたら、その単語が被覆された ($z_l = 1$) とするための制約である。式 2.6 は、ほぼ全ての作業者がニュース記事の一文目を要約に含めていたという観測に基づき、複数記事のうち、いずれかの記事の一文目を必ず要約に含めるという制約である。式 2.7 から式 2.9 は、ある二つの文が両方とも選択されたかどうかを変数 y に反映させるための制約である。

主題への関連度 $r_{(i,m)}$ は、次式で計算する。

$$r_{(i,m)} = \frac{1}{(\text{文}(i,m) \text{ の出現位置})} + (\text{文}(i,m) \text{ と文書集合の類似度}) \quad (2.1)$$

ここで、文 (i,m) は記事 i の m 番目の文を表す。文と文書集合の類似度と文間の類似度 $s_{(i,m)(j,n)}$ は、単語の Bag-of-Words のコサイン類似度で与える。

$$\begin{aligned} \max. \quad & \sum_{l=1}^L b_l z_l \times \left(\sum_{i=1}^I \sum_{m=1}^{M_i} r_{(i,m)} x_{(i,m)} - \sum_{(i,m) < (j,n)} s_{(i,m)(j,n)} y_{(i,m)(j,n)} \right) \\ \text{s.t.} \quad & \end{aligned} \quad (2.2)$$

$$\begin{aligned} \forall i, j, m, n, l : \quad & x_{(i,m)} \in \{0, 1\}, \quad y_{(i,m)(j,n)} \in \{0, 1\}, \\ & z_l \in \{0, 1\} \end{aligned} \quad (2.3)$$

$$\sum_{i=1}^I \sum_{m=1}^{M_i} c_{(i,m)} x_{(i,m)} \leq K \quad (2.4)$$

$$\forall l : \quad \sum_{i=1}^I \sum_{m=1}^{M_i} a_{(i,m)l} x_{(i,m)} \geq z_l \quad (2.5)$$

$$\sum_{i=1}^I x_{(i,1)} = 1 \quad (2.6)$$

$$\forall i, j, m, n : \quad y_{(i,m)(j,n)} - x_{(i,m)} \leq 0 \quad (2.7)$$

$$\forall i, j, m, n : \quad y_{(i,m)(j,n)} - x_{(j,n)} \leq 0 \quad (2.8)$$

$$\forall i, j, m, n : \quad x_{(i,m)} + x_{(j,n)} - y_{(i,m)(j,n)} \leq 1 \quad (2.9)$$

表 2.3: 変数の定義

$x_{(i,m)}$	文 (i, m) が選択されたかどうか
$y_{(i,m)(j,n)}$	文 (i, m) と文 (j, n) が両方選択されたかどうか
z_l	単語 l が被覆されたかどうか
K	最大要約長
$a_{(i,m)l}$	文 (i, m) が単語 l を含むかどうか
b_l	単語 l の重要度
$c_{(i,m)}$	文 (i, m) の長さ（要約長が文数の場合は1）
$r_{(i,m)}$	文 (i, m) の主題への関連度
$s_{(i,m)(j,n)}$	文 (i, m) と文 (j, n) の類似度
M_i	記事 i の文の数
I	記事の総数
L	単語の総数

2.4.3 単語の重要度

最大被覆モデルは、被覆された概念（ここでは、単語）の重要度の和が最大となるような文を選択するモデルである。そのため、単語の重要度の決め方が肝となる。従来、単語の重要度としては、頻度や TF-IDF が使われたり [Filatova 04]，ロジスティック回帰で推定した重要度 [Yih 07] などが用いられてきた。本研究では、単語の重要度を、RandomForest で推定した場合と BERT に基づくモデルで推定した場合について比較する。ここで、被覆する単語の品詞を動詞，名詞，形容詞，副詞に絞る（ただし，形式名詞，副詞的名詞，“する”は除いた）。

2.4.3.1 RandomForest を用いた単語重要度推定

RandomForest を用いて単語の重要度 b_l を推定する。学習のさせ方として以下の2通りを試した。一つは、作業者が対象単語を要約に含めたかどうかの2クラス分類問題として学習し、単語が要約に含まれる確率を単語の重みとしたもの（RFC）。もう一つは、複数の作業者が要約に含めた単語の重みが高くなるように回帰で学習し、推定したものである（RFR）。例えば、4人中3人が要約に含めた場合のその単語の重みを $3/4$ とし、4人中2人なら $2/4$ とする。素性としては、2.4.3.3 節のものを使用した。

2.4.3.2 BERT に基づく単語重要度推定モデル

BERT [Devlin 18] は, Transformer [Vaswani 17] の Encoder 部分をユニットとする双方向 Transformer モデルである. 文の単語をランダムにマスクし, そのマスクされた単語を予測する Masked Language Model と二つ文が隣接しているかどうかを予測する Next Sentence Prediction の2つのタスクで事前学習したモデルを転移学習させることで, 自然言語処理の様々なタスクで SOTA を達成し, 汎用的な言語表現を獲得できるモデルとして注目されている. 本研究では, BERT を抽出型の要約タスクに適用し, 事前学習の効果を検証する.

双方向 Transformer の BERT を基本構造として文脈を考慮して単語の重要度を推定するモデルを提案する (図 2.4). BERT から得られる単語の埋め込み表現に補助情報や作業員 ID を加え, 単語が非重要文, 重要文 a, 重要文 b に含まれる確率を softmax で推定する. 補助情報には, 2.4.3.3 節のものを使用した. 作業員 ID は, 複数の作業員のデータを混ぜて学習させるためのもので, 作業員情報を 1-of-k で区別して与える.

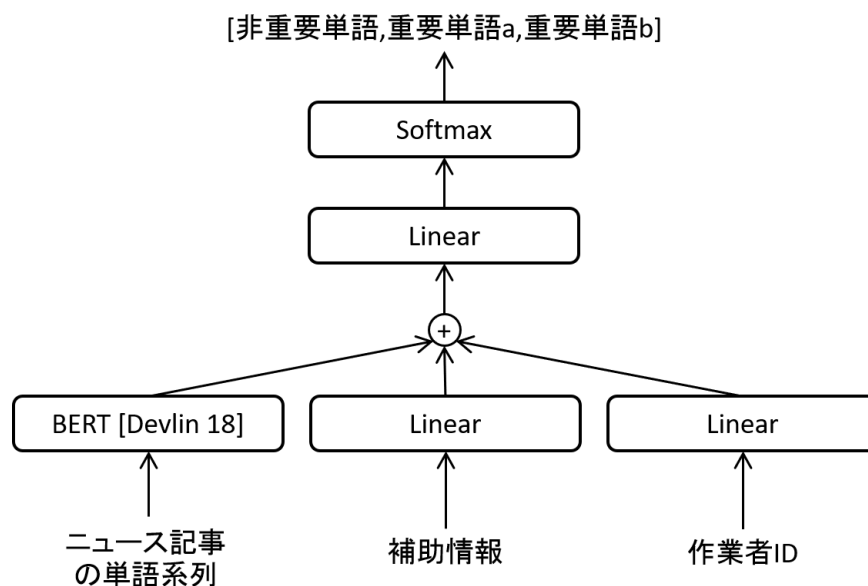


図 2.4: 単語重要度推定モデル

2.4.3.3 補助情報

JUMAN++¹ (Ver.1.02) の形態素情報 (品詞大分類, 品詞細分類, 活用形, 活用型, カテゴリ, ドメイン), 単語の TF, IDF, TF-IDF, タイトルに含まれるかどうか, 「」『』内の単語かどうか, 記事における文・段落の位置, KNP² (Ver.4.19) を適用して得られる IREX の 8 種類の固有表現クラス, 係り受けの種類, 係り受け木の深さ, 係り元の文節数, 文頭からの文節位置を用いた.

2.4.4 評価尺度 (ROUGE)

一般に, 良いシステム要約には, 人間が作成した正解要約に含まれる多くの語や句が出現するはずだと考えられる. 機械翻訳の分野でよく使われる評価尺度として BLEU (BiLingual Evaluation Understudy) [Papineni 02] というものがある. BLEU は精度重視の評価尺度であり, 再現率を重んじる要約の評価尺度としては適さない. そこで, Lin と Hovy は, BLEU を要約評価用に改良した評価尺度として, ROUGE を提案した [Lin 03]. ROUGE は以下の式で計算される.

$$\text{ROUGE}(i, j) = \exp \left(\sum_{n=i}^j \frac{1}{(j-i+1)} \log C_n \right) \quad (i \leq j) \quad (2.10)$$

ここで, C_n はシステム要約と正解要約の両方に共通して含まれる n -gram の数を正解要約中の n -gram の数で割った数値 (再現率) である.

Lin らは DUC2001 のデータを用いて ROUGE の有効性を検証した. その結果, $i = j = 1$ (ROUGE-1) のとき, 人間の評価者によるランクと最も相関があることが分かった. これは, 長い n -gram での評価は, 内容よりもむしろ文法の正しさを評価する傾向にあるためだと考えられる. 抜粋による要約では, 機械翻訳と異なり, 文法性の問題で困ることはないので, ROUGE-1 による評価で十分だと考えられる. しかしながら, 最近では, 抽象型の要約モデルも登場してきており [Rush 15, Chopra 16, Nallapati 16, Kikuchi 16, Paulus 18, Celikyilmaz 18], このようなモデルでは, 文法性の問題も発生するため, より高次の n -gram を考慮する必要があると考えられる. ROUGE には, 様々な種類が提案されている [Lin 04].

¹<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

²<http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

ROUGE-L (最長共通部分列一致)

ROUGE-L は、要約文を文字列の連続とみなして、システム要約と正解要約の最長共通部分列 (LCS: Longest Common Subsequence) を調べ、その長さを正解要約の長さで割った値をシステムの評価値とする。

ROUGE-W (重み付き部分文字列最長一致)

ROUGE-W は、ROUGE-L と同様、正解要約とシステム要約間の文字列の一致度により、評価値を算出するが、断片的に文字列が一致する場合よりも連続した一致により高い評価を与える点が異なる。

ROUGE-S (スキップバイグラム共起統計)

ROUGE-S は、正解要約から抽出されたスキップバイグラムと比較し、両者で一致するものの数を正解要約のスキップバイグラムの総数で割ったものである。スキップバイグラムとは、任意のギャップを許容した単語のペアを表す。

ROUGE-SN

ROUGE-SN は、スキップバイグラムを求める際、任意の2単語の組み合わせをとるのではなく、単語間のスキップ距離に上限 (N 単語) を設けたものである。

ROUGE-SU

ROUGE-SU は、ROUGE-S に unigram の一致を追加したものである。

ROUGE-WE

文字列の一致に基づく ROUGE では、同じ意味の単語であっても表記が異なれば誤りと判定されてしまう。Ng らは、ROUGE を文字列の一致ではなく、分散表現の類似度で計算する方法を提案した [Ng 15]。

本研究では、抜粋による要約であるため、文法が崩れる心配は少ないので、ROUGE-1 の値を重視するが、参考までに、ROUGE-2, ROUGE-3, ROUGE-S, ROUGE-SU, ROUGE-L の値も算出する。

2.4.5 Yahoo コーパスを用いた評価実験

2.4.5.1 実験設定

要約コーパスの5人の作業結果を用いて提案手法を評価する。提案手法では単語の重要度を学習する必要があるため、データセットを訓練セットとテストセットに分けて評価を行う。データセットは次の二つの観点で分割する。

- (1) 1人の作業結果を訓練セットとして、4人の作業結果をテストセットとする。
- (2) 4人の作業結果を訓練セットとして、1人の作業結果をテストセットとする。

また、100 トピックについて訓練セットとテストセットでトピックの重複がないように分割した。それぞれ、全ての組み合わせに関して ROUGE を計算し、その分布を箱ひげ図で描画する。そして、その分布が人間同士の比較に基づいて算出した ROUGE の分布のばらつき内に収まっているかどうかでモデルの良し悪しを判断する。また、MCKP 単体と MCKP+Rel の比較も行う。なお、最大要約長 K は、各トピックで作業者が選択した文の数とする。

「単語の重要度」の節で、2通りの重要度の与え方を紹介したが、(1)のときは、作業者が対象単語を要約に含めたかどうかの2クラス分類問題として RandomForest を学習する (RFC)。(2)のときは、複数の作業者が要約に含めた単語の重みが高くなるように回帰で学習する (RFR)。4人のデータを訓練セットとしているため、4人中 n 人が要約に含めた単語の重みを $n/4$ として学習する。

2.4.5.2 実験結果

1人のデータで訓練し、4人のデータでテストしたときの結果を図2.5に示す。4人のデータで訓練し、1人のデータでテストしたときの結果を図2.5に示す。モデル同士の比較では、MCKP+Relの方がわずかに良い結果を示した。データセットの分割方法に関して結果を比較すると、4人のデータで学習したときの方が良い結果を示していることが分かる。これは、複数人のデータを用いることで、モデルの汎化性能が向上したためだと考えられる。また、複数人のデータを考慮した回帰による単語の重み付けの効果もあると考えられる。

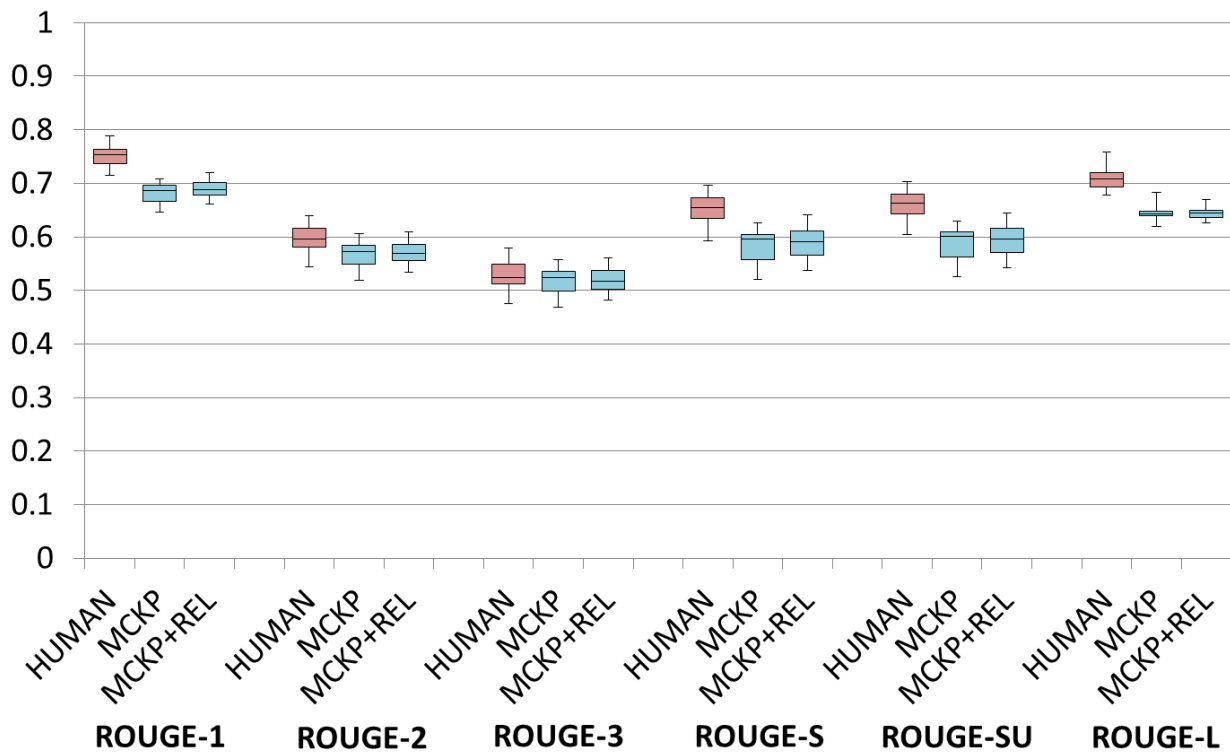


図 2.5: 訓練 1 人-テスト 4 人のときの結果

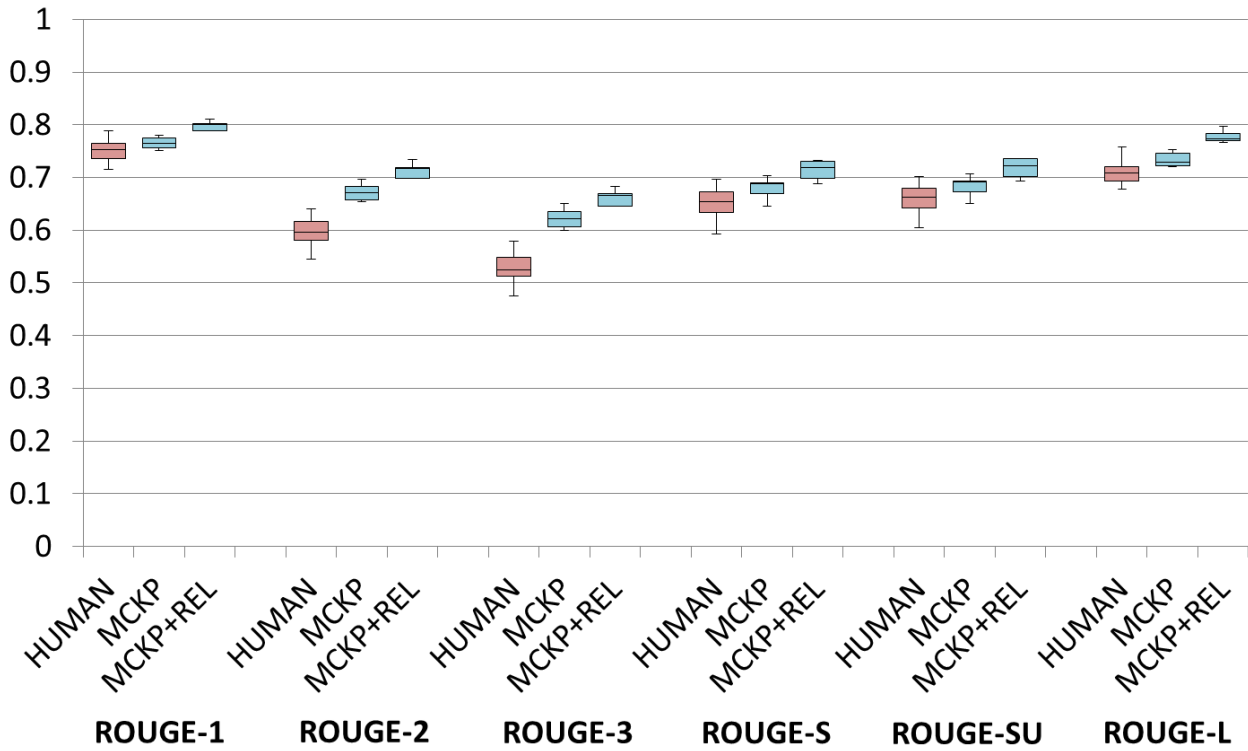


図 2.6: 訓練 4 人-テスト 1 人のときの結果

2.4.5.3 計算量

最大被覆問題はNP困難な問題である。Filatovaらは単語を概念とし、重要度としてTF-IDFを与え、貪欲法で解いた[Filatova 04]。Yihらは文書集合での出現頻度に比例する重みによる単語の重み付けと訓練データで学習したロジスティック回帰の確率値を利用した単語の重み付けを行い、スタック・デコーディングを用いて最大被覆問題を解いた[Yih 07]。高村らは最大被覆問題を整数計画問題として厳密に定式化し、様々なデコーディング方法の比較を行った[Takamura 09]。これらの研究のように計算量の問題から、一般に、整数計画問題の解法としては、近似解を得るアルゴリズムが用いられるが、本研究では、数十文から数文を選択する程度の比較的小規模な要約問題であったため、分枝限定法により厳密解を得ることが可能であった。しかしながら、実際のシステムに適用するにはスケーラビリティの問題があるため、最大被覆モデル単体を使用する。予備実験で、MCKP+Relを貪欲法で解くよりも、MCKP単体を分枝限定法で解いた方が高いROUGEを示すことを確認した。

MCKP+Relの要約長ごとの入力文数に対する計算時間を図2.7に示す。抽出文数が4になると急激に計算時間が増加することが分かる。次に、MCKP単体、Rel単体での入力文数が20文固定のときの要約長に対する計算時間を図2.8に示す。MCKP単体の場合、要約長が増えても計算時間はそれほど急激には増加しないことが分かる。

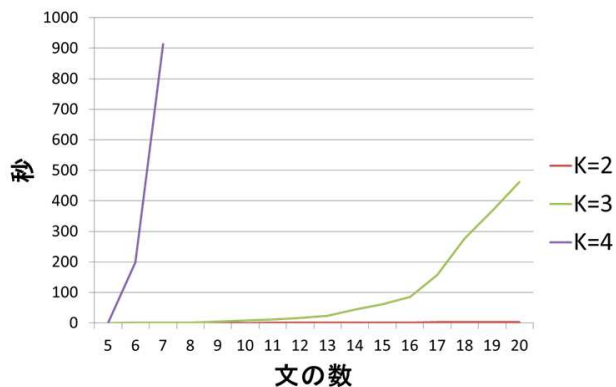


図 2.7: MCKP+Rel の要約長ごとの入力文数に対する計算時間

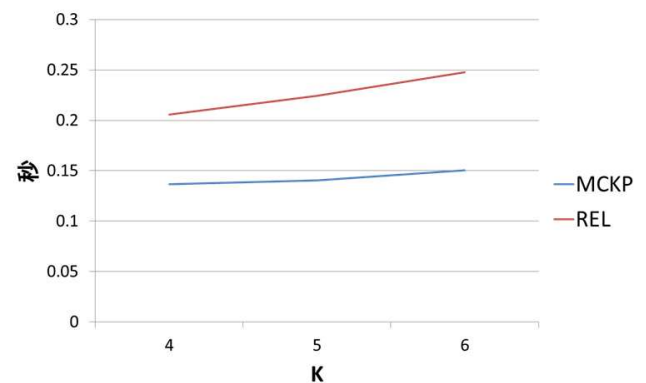


図 2.8: MCKP 単体, Rel 単体での入力文数が20文固定のときの要約長に対する計算時間

2.4.6 日経コーパスを用いた評価実験

2.4.6.1 BERT の事前学習

日経新聞の200476個のニュース記事から段落を超えないように隣接文ペアを重複なく抽出した。このうち、700000文ペアを訓練セット、37094文ペアを開発セットとして、Masked Language Model と Next Sentence Prediction の2タスクでBERTの事前学習を行った。語彙には訓練セットにおいて頻度が7以上であった63272語を用いた。モデルのパラメータは、Transformerのブロック数を $L = 8$ 、隠れ層の次元を $H = 256$ 、self-attentionのヘッド数を $A = 8$ に設定した。

2.4.6.2 実験設定

日経コーパスを用いて、提案モデルによって推定した単語重要度を用いたときの重要文抽出の性能を評価する。作業員*i*のデータ90個を訓練、作業員*j*のデータ10個をテスト($i \neq j$)となるように分割し、10分割交差検定で評価した。評価尺度にはROUGE-1を用いた。抽出する文の数 K は作業員が選択した文の数とした。整数計画問題を解くための最適化アルゴリズムには分枝限定法を用いた。被覆する単語の品詞は、名詞、動詞、形容詞、副詞に限定した。

事前学習なし、補助情報なしで単一作業員のデータを用いて学習したモデルをSEM1、これに事前学習を行ったモデルをSEM1_pre、さらに補助情報を加えて学習したモデルをSEM1+_pre、作業員2人のデータを用いて学習したモデルをSEM2+_preとする。

比較手法として、2.4.3.3節で説明した補助情報を素性として学習したRandomForestにより単語の重要度を与えたときの結果(RF)、作業員の一方をシステム出力、他方を正解としてROUGE-1を算出した結果(Human)、ランダムに文を抽出したときの結果(Random)、先頭から K 文抽出したときの結果(LEAD)と比較を行った。

2.4.6.3 実験結果

重要文aを抽出する実験の結果を図2.9に、重要文bを抽出する実験の結果を図2.10に示す。重要文a、重要文bのいずれの抽出に関してもRandomForestで重要度を与えた場合よりも提案手法の方が良い結果を示した。また、事前学習なしのモデル(SEM1)よりも事前学習を行ったモデル(SEM1_pre)の方が良く、さらに補助情報を加えて学習することで(SEM1+_pre)、複数人のデータを活用することで(SEM2+_pre)性能が向上することが分かった。

重要文aの抽出に関しては、LEAD法が高い性能を示した。これは重要文aが文の初めの方から選ばれていることに起因する(図2.2)。一方で、重要文bの抽出においては、LEAD法はRandom以下の性能まで落ち込むことが分かった。

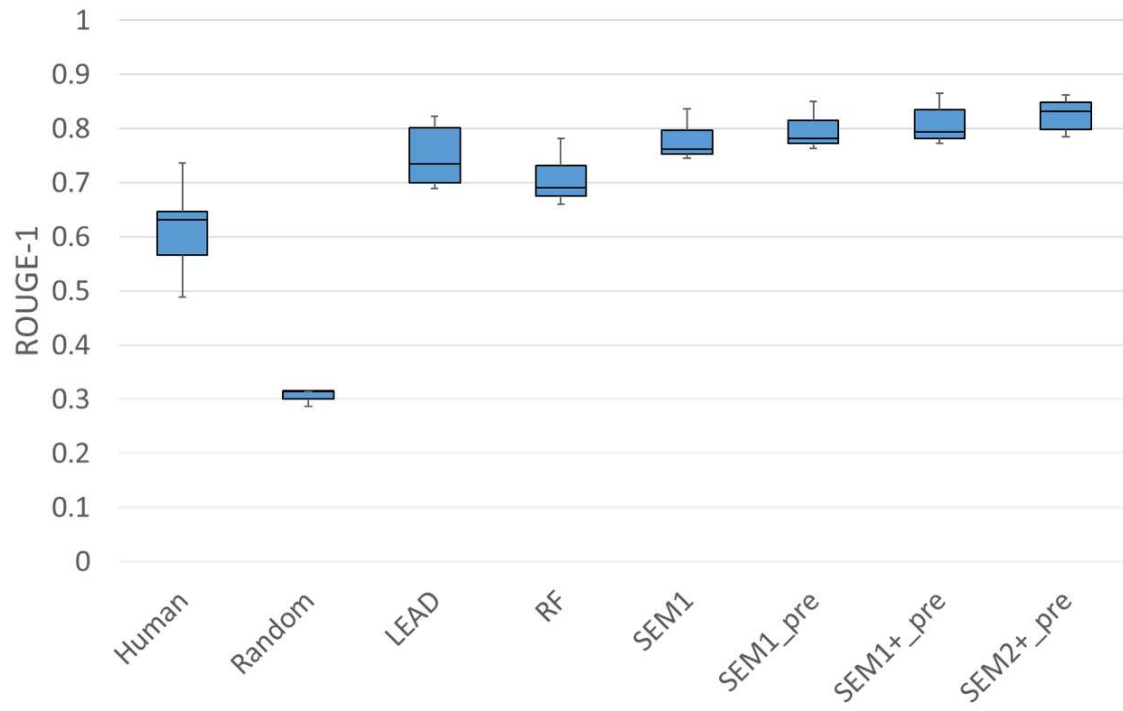


図 2.9: 重要文 a 抽出の実験結果

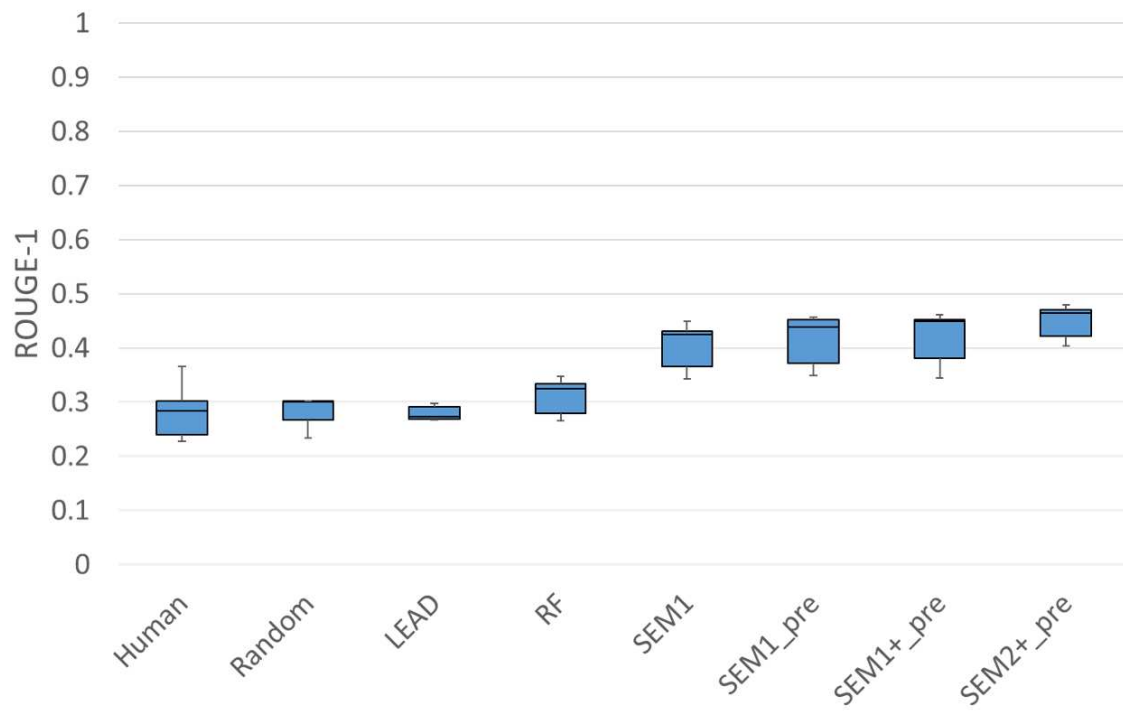


図 2.10: 重要文 b 抽出の実験結果

2.5 整列

重要文抽出で抽出した文の提示順序を決定する。単一文書要約の場合、記事における文の出現順序をそのまま利用できるが、複数文書要約の場合、異なる記事から文が選択される可能性があるため、要約内での文の提示順序を決める必要がある。

2.5.1 関連研究

文の整列問題は、主にテキスト生成の分野で取り組まれてきたが、近年では、自動要約の分野でも重要な課題として認識されている。岡崎らは記事が書かれた時間情報を利用して並び替える手法を提案した [Okazaki 04]。Lapata は文中での動詞や名詞とその他の係り受け関係を考慮した統計的なモデルに基づいて文の順序を決定する手法を提案した [Lapata 03]。最近では、RNN 言語モデルを用いて文の整列を試みる研究も行われており、Lin らは文間の結束性を考慮した RNN 言語モデル (Hierarchical Recurrent Neural Network Language Model) を提案し、HRNNLM が文整列タスクにおいて、最大エントロピー法に基づく手法や再帰ニューラルネットワークに基づく手法よりも優れていることを示した [Lin 15]。

2.5.2 複数文書要約における整列アルゴリズム

複数記事から抽出された文の整列手法を提案する。要約コーパスにおいて同じ記事から選ばれた文の順序関係が保たれていたことから、ある記事の文が別の記事のどの文に該当するかをコサイン類似度に基づいて計算し、最もスコアが高くなる順序を採用するという方法をとった (図 2.11)。



図 2.11: 整列の考え方

$$sequence = \max_{seq \in S} Score(seq) \quad (2.11)$$

$$Score(seq) = \sum_{(i,m) \rightarrow (j,n) \in seq} f((i,m), (j,n)) \quad (2.12)$$

$$f((i,m), (j,n)) = \quad (2.13)$$

$$\begin{cases} g(\sum_{l>m} \text{sim}((i,l), (j,n)), \sum_{l<n} \text{sim}((i,m), (j,l))) & (2.14) \end{cases}$$

$$\begin{cases} g(\text{sim}((i,m), \text{before}(j,n)), \text{sim}(\text{after}(i,m), (j,n))) & (2.15) \end{cases}$$

ここで、 S は文の順序候補の集合、 $f((i,m), (j,n))$ は文 (i,m) が文 (j,n) よりも前に来うるスコア、 g, h は \sum または \max 、 sim は BoW のコサイン類似度を表す。また、 $\text{before}(j,n)$ は記事 j の文 n よりも前に出現する文の BoW を表し、 $\text{after}(i,m)$ は記事 i の文 m よりも後ろに出現する文の BoW を表す。

2.5.3 Yahoo コーパスを用いた評価実験

要約コーパスには単一記事からなるトピックも含まれるため、単一記事要約の場合は記事の出現順番をそのまま使用した。作業者5人の結果を正解として、正解率を計算し、箱ひげ図に描画した結果を図2.12に示す。人間（総合）は、異なる2人の組み合わせに関して、重要文抽出で同じ文集合が選ばれたトピックのうち、選択した全ての文が同じ順番に並び替えられたものの割合を表す。ここで、人間同士の比較に基づいて整列の正解率を算出した結果の詳細を表2.4に示す。表から分かる通り、重要文抽出で同じ文集合が選ばれるトピックの数がそもそも少ないので、ばらつきが大きい。特に複数記事の数が少ない。そのため、単一記事と複数記事を区別せず、総合の正解率のみを描画した。システムの正解率の詳細を表2.5に示す。

単一記事の場合、そのままの順番でも85%程度の正解率を実現しているが、中には順番が入れ替わっているトピックもある。例えば、図2.13の文章のように今後の展望など、会話の締めとなる内容が文章の中盤に記載されている場合、順番が入れ替わることがある。

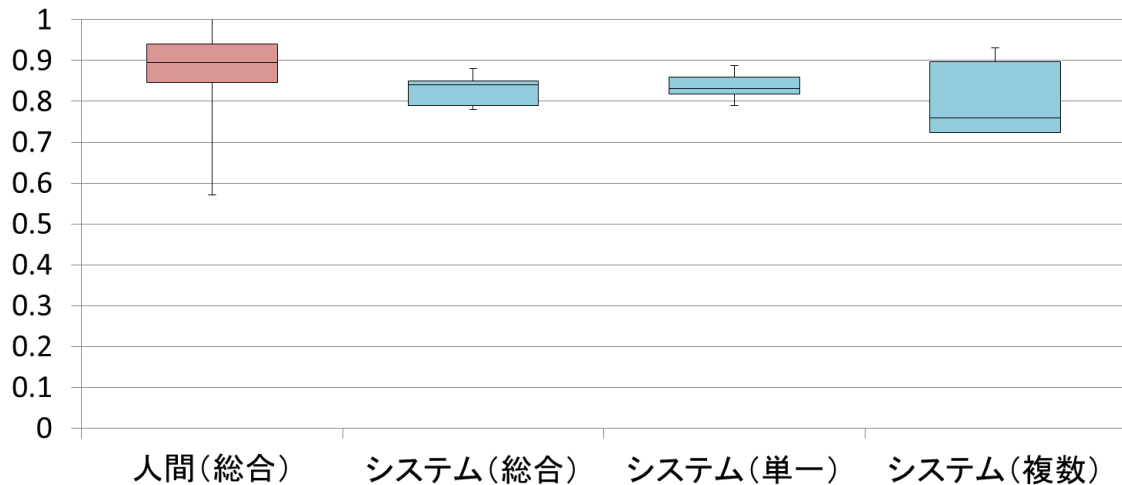


図 2.12: 整列結果の正解率

表 2.4: 人間同士の比較結果の詳細

作業者の組み合わせ	単一記事における正解率	複数記事における正解率	総合正解率
作業者 1 - 作業者 2	1.00 (11/11)	0.00 (0/0)	1.00 (11/11)
作業者 1 - 作業者 3	0.50 (3/6)	1.00 (1/1)	0.57 (4/7)
作業者 1 - 作業者 4	0.71 (5/7)	1.00 (1/1)	0.75 (6/8)
作業者 1 - 作業者 5	0.89 (8/9)	0.00 (0/0)	0.89 (8/9)
作業者 2 - 作業者 3	0.88 (7/8)	1.00 (2/2)	0.90 (9/10)
作業者 2 - 作業者 4	0.93 (14/15)	0.50 (1/2)	0.88 (15/17)
作業者 2 - 作業者 5	0.93 (13/14)	1.00 (3/3)	0.94 (16/17)
作業者 3 - 作業者 4	1.00 (13/13)	0.00 (0/0)	1.00 (13/13)
作業者 3 - 作業者 5	0.91 (10/11)	0.00 (0/1)	0.83 (10/12)
作業者 4 - 作業者 5	1.00 (15/15)	0.00 (0/1)	0.94 (15/16)

表 2.5: システムの正解率の詳細

	単一記事における正解率	複数記事における正解率	総合正解率
作業者 1	0.83 (59/71)	0.90 (26/29)	0.85 (85/100)
作業者 2	0.82 (58/71)	0.72 (21/29)	0.79 (79/100)
作業者 3	0.89 (63/71)	0.72 (21/29)	0.84 (84/100)
作業者 4	0.86 (61/71)	0.93 (27/29)	0.88 (88/100)
作業者 5	0.79 (56/71)	0.76 (22/29)	0.78 (78/100)

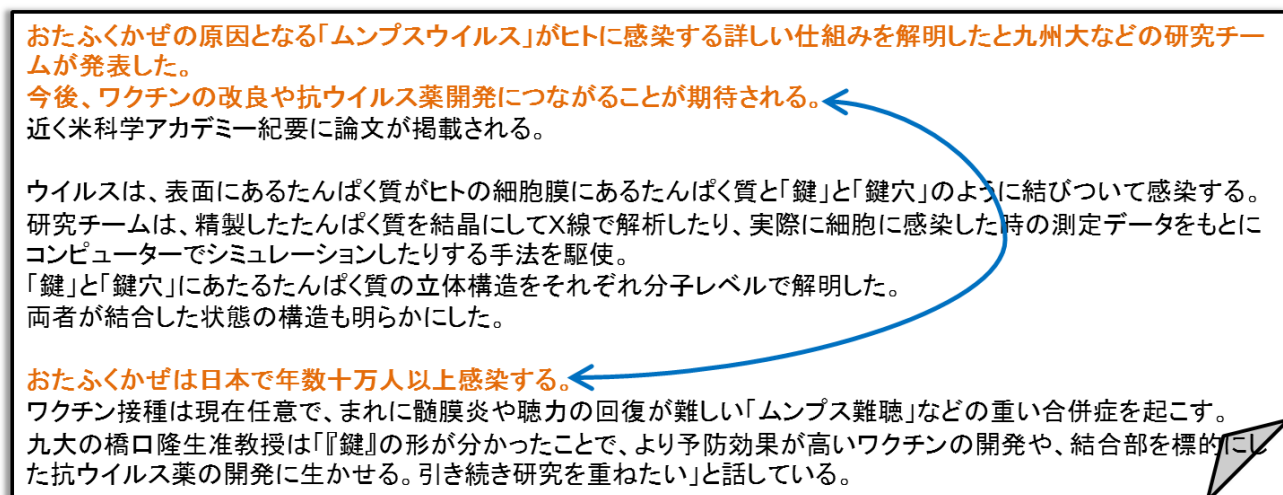


図 2.13: 単一記事で順序が入れ替えられたトピックの例

2.6 文圧縮

視覚メディアによる情報伝達を前提として書かれたニュース記事の一文一文には、たくさんの情報が埋め込まれている。もしその内容を会話で伝えるのであれば、話し手は送受信の同時性に起因する聞き手の負担を考え、極力情報量を抑え、分かりやすい表現で伝えなければならない [畠 87]。そこで、情報を削減する手段として要約技術、特に、文の内容を短く縮約する手段として文圧縮技術の利用が考えられる。

Differentiable Neural Computers (DNC) をはじめとするいくつかの系列ラベリングのモデルを Yahoo コーパスで比較する。また、BERT に基づく文圧縮モデルを提案し、文法性を保証する枠組みとして野本の手法 [Nomoto 08] を導入する。そして、このモデルを日経コーパスで評価する。

2.6.1 関連研究

文圧縮の技術は、文字放送や字幕作成、限られた表示リソースしか持たない携帯端末でのテキスト表示を支援する技術として注目されている。字幕作成では、表示する機器の大きさや人間の読むスピードに制限があることから、体言止めや漢字熟語を多用した表現への書き換えなどが行われる [江原 02]。一方、本研究では、会話で話すための話し言葉への書き換えを目的としているため、このような書き換えは行わず、ここでは、不要な文節の削除のみを行う。なお、表現の書き換えについては後段の口語化処理で行う (3章参照)。

文圧縮は文から情報を削り、文自体を短く縮約するタスクである。統計的な文圧縮の手法とし

て, Knight と Marcu は, Noisy-Channel モデルと決定木による文圧縮手法を提案した [Knight 02]. Clarke と Lapata は, 文圧縮問題を整数計画問題として定式化し, 制限長内で重要度を最大化するような単語群を含めるような文を生成する手法を提案した [Clarke 08]. 野本は, 係り受け木の刈り込みにより要約文候補を生成し, CRF を用いて選別する手法を提案した [Nomoto 08]. 最近では, Encoder-Decoder モデルにより文全体のエンコード結果を考慮して系列ラベリングの要領で単語を選択するモデル [Filippova 15] や, 要約モデルにより要約文を生成する抽象型の手法 [Rush 15, Chopra 16, Nallapati 16, Kikuchi 16] がある. 抽象型の手法には, 文書の内容と無関係な単語を生成してしまう問題や文書の内容と矛盾する要約を生成してしまう問題がある. このような誤生成問題に取り組んだ研究はあるものの [See 17, Suzuki 17, 清野 18], 実用的なレベルには至っていない. また, 学習には膨大なデータを必要とする. そこで, 本研究では, 文圧縮の問題を文節単位の系列ラベリングの問題として定式化する.

2.6.2 系列ラベリングモデル

本研究では, 文圧縮の問題を系列ラベリングの問題として捉え, 文節単位で発話に含めるか否かを決定する. 発話に含める文節にはラベル “I” を付与し, 発話に含めない文節にはラベル “O” を付与する. ここでは, モデルとして CRF と双方向 LSTM, Pointer Networks, DNC を用いて比較を行った. これらのモデルの学習には, 素性として, 文節を構成する単語の BoW と 2.6.4 で説明する補助情報を用いた.

2.6.2.1 Conditional Random Fields

1 つ目のモデルは Conditional Random Fields (CRF) である. CRF は自然言語処理において形態素解析 [Kudo 04, Kumar 14] や固有表現抽出 [Lafferty 01, Rahman 16] などの系列ラベリングの問題で幅広く使われている.

CRF のツールとして CRF++³を使用した. 過学習を防ぐために L2 正則化を行い, 正則化項の係数は 1.0 とした.

³<https://taku910.github.io/crfpp/>

2.6.2.2 Bidirectional LSTM

2つ目のモデルは双方向 LSTM (BiLSTM) である。双方向 LSTM の構造を図 2.14 に示す。双方向 LSTM は、文節の素性ベクトル \mathbf{x} を入力として、埋め込み層を経て、順方向と逆方向の隠れ層に伝搬し、同時刻の順方向の隠れ層と逆方向の隠れ層の結合を出力層に伝搬する構造となっている。出力層では softmax を計算して、各ラベルの出力確率 \mathbf{y} を求める。

実験では、埋め込み層と隠れ層を 200 次元に設定した。最適化アルゴリズムには Adam を用いた。

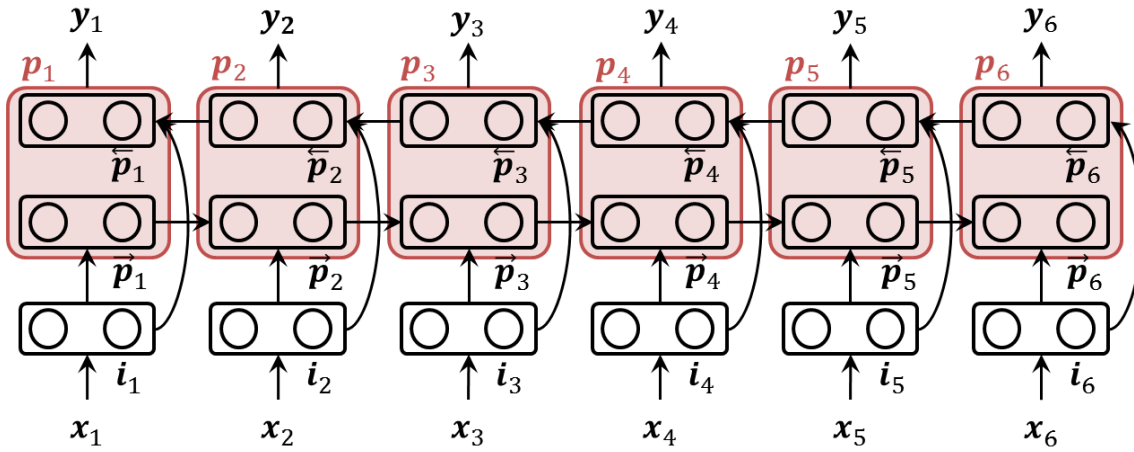


図 2.14: Bidirectional LSTM

$$\mathbf{i}_n = \tanh(W_{xi} \cdot \mathbf{x}_n) \quad (2.16)$$

$$\vec{\mathbf{p}}_n = \text{LSTM}\left(\vec{W}_{ip} \cdot \mathbf{i}_n + \vec{W}_{pp} \cdot \mathbf{p}_{n-1}\right) \quad (2.17)$$

$$\overleftarrow{\mathbf{p}}_n = \text{LSTM}\left(\overleftarrow{W}_{ip} \cdot \mathbf{i}_n + \overleftarrow{W}_{pp} \cdot \mathbf{p}_{n+1}\right) \quad (2.18)$$

$$\mathbf{p}_n = [\vec{\mathbf{p}}_n || \overleftarrow{\mathbf{p}}_n] \quad (2.19)$$

$$\mathbf{y}_n = \text{softmax}(W_p \cdot \mathbf{p}_n) \quad (2.20)$$

2.6.2.3 Pointer Networks

3つ目のモデルは Pointer Networks (PN) である． Pointer Networks は，入力の位置に関するポインタを出力とする Encoder-Decoder モデルである． Vinyals らはこのモデルを組合せ最適化問題に適用し，その近似解が得られることを示した [Vinyals 15a]． 現在研究が進められている生成型要約の手法においても，要約では原文の単語が使われやすいことから， Pointer Networks のような機構をモデルに導入して，出力辞書にない低頻度な単語を入力文章から参照しようという試みがなされている [Gulcehre 16, See 17, Paulus 18, Celikyilmaz 18]．

オリジナルの Pointer Networks では， Sutskever らのモデル [Sutskever 14] と同様， Encoder から Decoder への入力は一時的な構造となっている（図 2.15）． そこで， Decoder によるより直接的な入力参照を可能とするために，本研究では Bahdanau らのモデル [Bahdanau 15] のように Encoder を双方向 LSTM にして attention 機構を導入した（図 2.16）．

実験では， Encoder と Decoder の隠れ層ともに 200 次元に設定した．

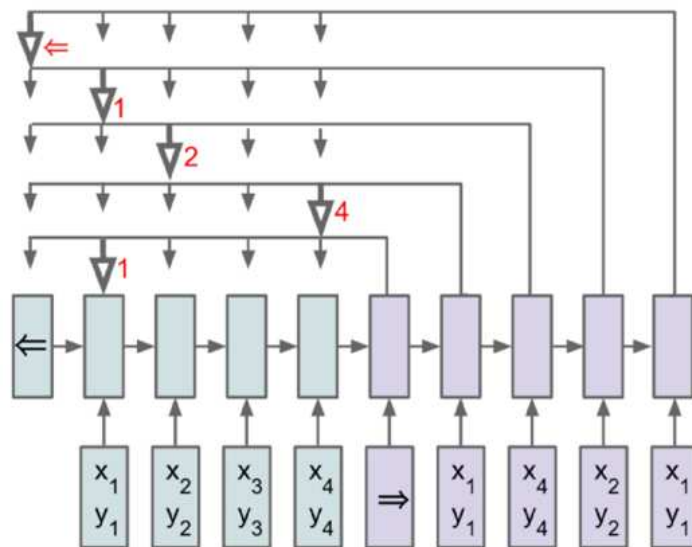


図 2.15: オリジナルの Pointer Networks [Vinyals 15a]

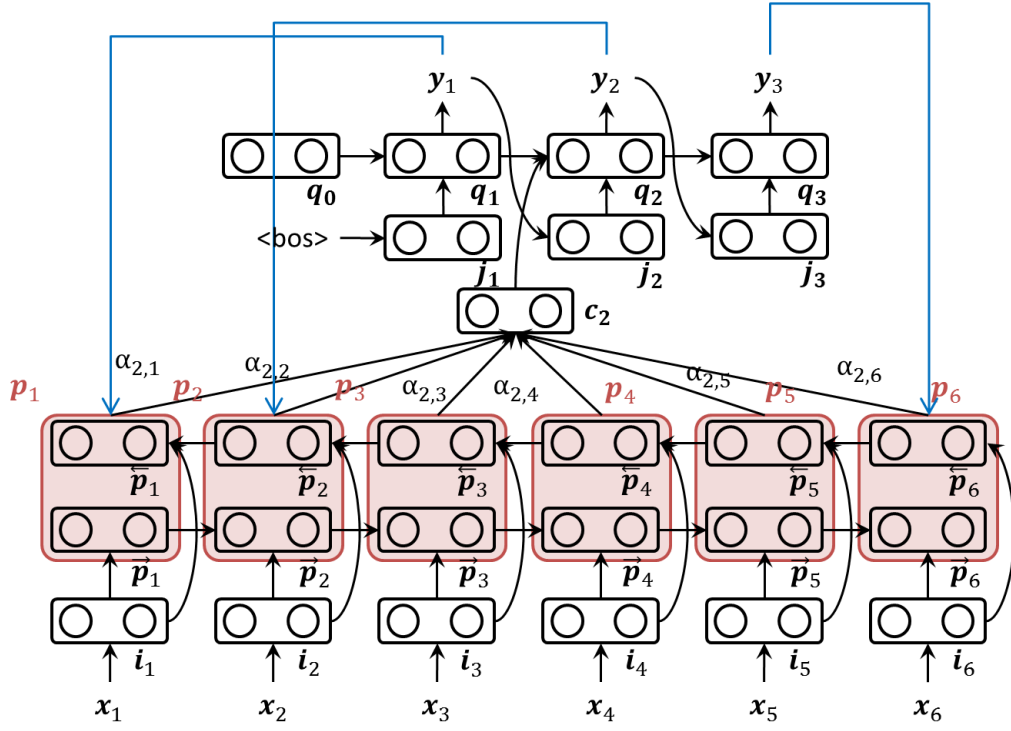


図 2.16: Pointer Networks with Attention

$$\mathbf{i}_n = \tanh(W_{xi} \cdot \mathbf{x}_n) \quad (2.21)$$

$$\vec{\mathbf{p}}_n = \text{LSTM}(\vec{W}_{ip} \cdot \mathbf{i}_n + \vec{W}_{pp} \cdot \mathbf{p}_{n-1}) \quad (2.22)$$

$$\overleftarrow{\mathbf{p}}_n = \text{LSTM}(\overleftarrow{W}_{ip} \cdot \mathbf{i}_n + \overleftarrow{W}_{pp} \cdot \mathbf{p}_{n+1}) \quad (2.23)$$

$$\mathbf{p}_n = [\vec{\mathbf{p}}_n || \overleftarrow{\mathbf{p}}_n] \quad (2.24)$$

$$u_{mn} = \mathbf{v}^\top \tanh(W_p \cdot \mathbf{p}_n + W_q \cdot \mathbf{q}_{m-1}) \quad n \in (1, \dots, N) \quad (2.25)$$

$$\alpha_{mn} = \text{softmax}(u_{mn}) \quad n \in (1, \dots, N) \quad (2.26)$$

$$\mathbf{c}_m = \sum_{n=1}^N \alpha_{mn} \mathbf{p}_n \quad (2.27)$$

$$\bar{\mathbf{p}} = \frac{1}{N} \sum_{n=1}^N \mathbf{p}_n \quad (2.28)$$

$$\mathbf{q}_0 = \tanh(W_{q_0} \cdot \bar{\mathbf{p}}) \quad (2.29)$$

$$\hat{n} = \text{argmax}(\mathbf{y}_{m-1}) \quad (2.30)$$

$$\mathbf{j}_{\hat{n}} = \tanh(W_{yj} \cdot \mathbf{x}_{\hat{n}}) \quad (2.31)$$

$$\mathbf{q}_m = \text{LSTM}(W_c \cdot \mathbf{c}_m + W_{yq} \cdot \mathbf{j}_{\hat{n}} + W_{qq} \cdot \mathbf{q}_{m-1}) \quad (2.32)$$

$$u'_{mn} = \mathbf{v}'^\top \tanh(W'_p \cdot \mathbf{p}_n + W'_q \cdot \mathbf{q}_m) \quad n \in (1, \dots, N) \quad (2.33)$$

$$y_{mn} = \text{softmax}(u'_{mn}) \quad n \in (1, \dots, N) \quad (2.34)$$

2.6.2.4 Differentiable Neural Computers

4つ目のモデルは Differentiable Neural Computers (DNC) である。DNC は外部メモリを有するニューラルネットワークである [Graves 16] (図 2.17)。DNC のメモリには、読み込みヘッドと書き込みヘッドがついており、ヘッドの位置を動かすことで、ヘッドの位置の情報を読み込んだり、書き込んだりすることができる。これらのヘッドを制御するのが CPU に相当するコントローラである。コントローラには入力ベクトルの他、読み込みヘッドで前時刻に読み込んだ読み込みベクトルも一緒に入力として与えられる。また、出力として出力ベクトルの他、メモリの書き込みに使われる様々なパラメータが出力される。

これまでのニューラルネットワークでは、直感的な「速い思考」しかできなかったのに対し、DNC の登場で経済学者のダニエル・カーネマンが言うところの論理的な「遅い思考」[カーネマン 14] が可能になるのではないかと期待されている。

コントローラとしては様々なタイプのニューラルネットワークを用いて設計することができるが、ここでは LSTM を用いた。実験では、ハイパーパラメータであるメモリスロットの数と次元を 100 とし、読み込みヘッドの数を 5 に設定した。

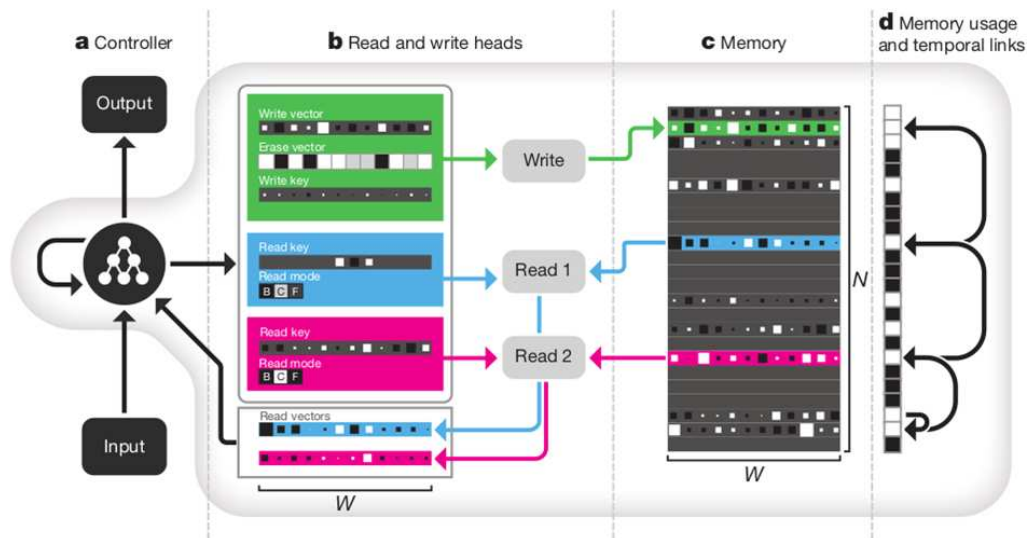


図 2.17: Differentiable Neural Computers [Graves 16]

2.6.3 BERT に基づく文圧縮モデル

双方向 Transformer の BERT に基づく文圧縮モデルを図 2.18 に示す．BERT で得られた単語の埋め込み表現を self-attention [Lin 17] によって文節ごとに組み上げている点が単語重要度推定モデル（図 2.4）と異なる．系列ラベリングにおける IOBES ラベルの基準で，文節を省くか含めるかの確率を softmax で推定する．補助情報には，2.6.4 節のものを使用した．作業 ID は，複数の作業者のデータを混ぜて学習させるためのもので，作業者情報を 1-of-k で区別して与える．

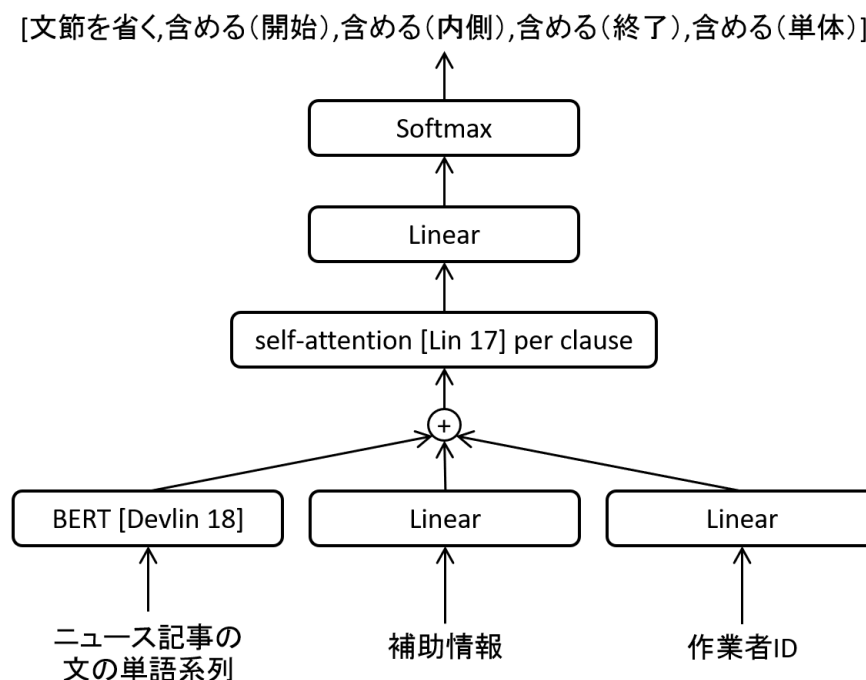


図 2.18: 文圧縮モデル

2.6.4 補助情報

JUMAN++⁴ (Ver.1.02) の形態素情報（品詞大分類，品詞細分類，活用形，活用型，カテゴリ，ドメイン），単語の TF，IDF，TF-IDF，タイトルに含まれるかどうか，「」『』内の単語かどうか，KNP⁵ (Ver.4.19) を適用して得られる IREX の 8 種類の固有表現クラス，係り受けの種類，係り受け木の深さ，係り元の文節数，文頭からの文節位置を用いた．

⁴<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

⁵<http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

2.6.5 Yahoo コーパスを用いた評価実験

2.6.5.1 実験設定

要約コーパスの5人の作業結果を用いて各モデルの評価を行う。データセットは次の二つの観点で、訓練セットとテストセットに分割する。

- (1) 1人の作業結果を訓練セットとして、4人の作業結果をテストセットとする。
- (2) 4人の作業結果を訓練セットとして、1人の作業結果をテストセットとする。

会話による情報伝達を目的とした文圧縮では、いかに情報を削れるかが重要であるため、評価尺度としてF値を用いる。F値は精度と再現率の調和平均であり、両者のバランスが重要となる。実験ではラベルIとラベルOのF値を計算し、それぞれを箱ひげ図で描画する。

モデルの良し悪しは、人間同士の比較結果を参考に判断する。人間同士の比較では、異なる2人の組み合わせに対して、一方をシステム出力、もう一方を正解データと見なしてF値を計算する。ここで、異なる2人の文圧縮結果を比較することになるが、文圧縮作業は重要文抽出作業で選択した文に対して行われるため、全ての文圧縮結果が比較できるわけではない。そこで、2人が重要文抽出作業で共通して選んだ文に対してのみ比較を行う。異なる2者間で共通して選ばれた文の数を表2.6に示す。重要文抽出作業で、各作業者は約300文を選択しており、表から、そのうち、およそ半数が異なる2者間で一致して選ばれていることが分かる。

表 2.6: 異なる2者間で共通して選ばれた文の数

作業者の組み合わせ	共通して選ばれた文の数
作業者1 - 作業者2	158
作業者1 - 作業者3	153
作業者1 - 作業者4	159
作業者1 - 作業者5	160
作業者2 - 作業者3	164
作業者2 - 作業者4	181
作業者2 - 作業者5	173
作業者3 - 作業者4	170
作業者3 - 作業者5	162
作業者4 - 作業者5	172

作業者ごとに対象としている文が異なるということを踏まえ、システムは以下の3通りで評価する。

訓練1人 - テスト 4人(全データ使用)

1人の作業結果を訓練セットとして、4人の作業結果をテストセットとする。テストで使用する作業者の結果は全て使用して評価する。

訓練1人 - テスト 4人(共通データのみ使用)

1人の作業結果を訓練セットとして、4人の作業結果をテストセットとする。ただし、テストで使用する作業者の結果のうち、訓練で使った作業者と共通の文のみを使用して評価する。これは、人間同士の比較と同じ条件である。

訓練4人 - テスト 1人(全データ使用)

4人の作業結果を訓練セットとして、1人の作業結果をテストセットとする。テストで使用する作業者の結果は全て使用して評価する。5人に共通して選ばれる文はほとんどないため、全て使用した場合のみの評価とする。

2.6.5.2 実験結果

「実験設定」の節で説明した3通りの方法で評価した結果を図2.19、図2.20、図2.21に示す。モデル間での比較では、DNCが最もよく、続いて Pointer Networks (PN) と双方向LSTM (BiLSTM) がほぼ同程度の性能を示した。一方、CRFはこれらに比べ、大きく下回る結果であった。これは、LSTMや外部メモリを使用したモデルでは、文節間の長期的な依存関係を捉えられているのに対し、高々数次のマルコフ性しか仮定していないCRFでは、こういった長期的な関係性を捉えきれていないためだと考えられる。また、ラベルIとラベルOの結果を比較すると、表2.1の要約率から分かる通り、ラベルOの方がラベルIよりもデータ数が少ないため、誤りやすい傾向にある。特に、CRFではラベルの偏りの影響を受けやすく、OをIと間違えることが多い。そのため、文圧縮の結果が要約率の低いものになってしまう。会話による情報伝達では、いかに情報を削れるかが重要であるため、CRFは目的にそぐわない。実際、CRFの結果を見てみると、長い文ほど誤りやすいことが確認された。一方、DNCでは長い文でも削るべき文節を正しく捉えられており、外部メモリの有効性が伺える。

人間同士の比較結果と比べると、同一条件での結果（訓練1人 - テスト4人（共通データのみ使用））において、CRF以外は人間同士の比較結果を上回る結果を示した。ここで、図2.19と図2.20を比較して、テストで全データを使用したときよりも、共通して選ばれた文のみを用いたときの方が高いF値を示していることが分かる。共通して選ばれた文というのは、ニュースの核となる情報であることから、モデルは揺らぎの少ない部分（核となる情報）にチューニ

ングされた学習をしていることが分かる．人間のデータには，そもそも嗜好性の違いによる揺らぎが存在しており，揺らぎの大きい人間同士の結果を比較した場合よりも，揺らぎの少ない部分（核となる情報）にチューニングして学習されたモデルの方が全体に対する平均正答率が高かったため，人間同士の比較結果を上回る結果を示したと考えられる．また，4人のデータで訓練したときの結果が1人のデータで訓練したときの結果よりも良かった理由としては，複数人のデータを用いたことで，モデルの汎化性能が向上したためだと考えられる．

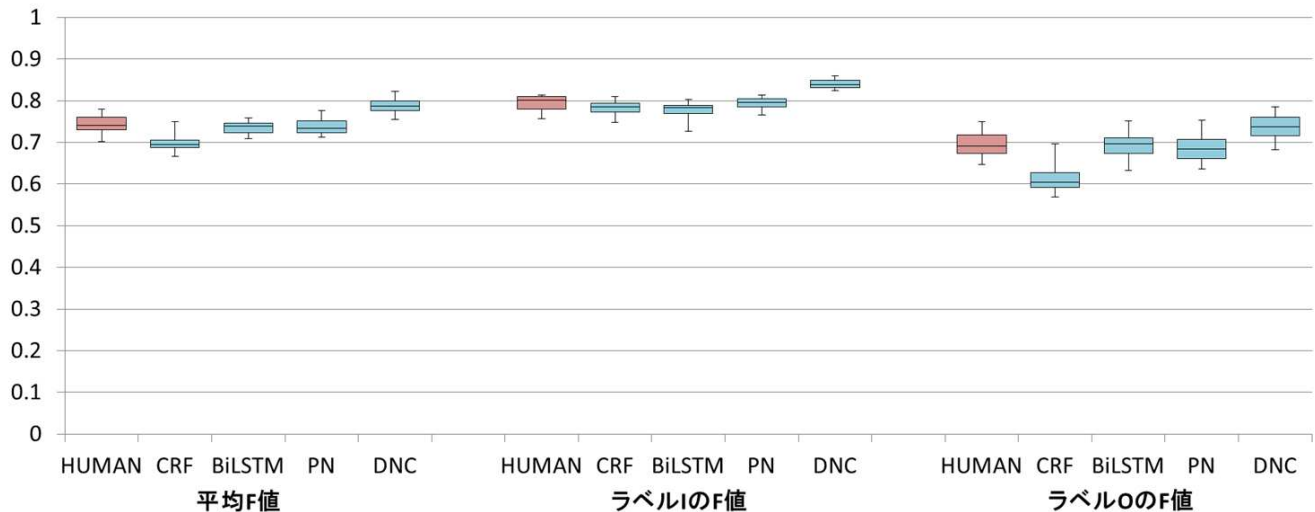


図 2.19: 訓練1人 - テスト4人（全データ使用）

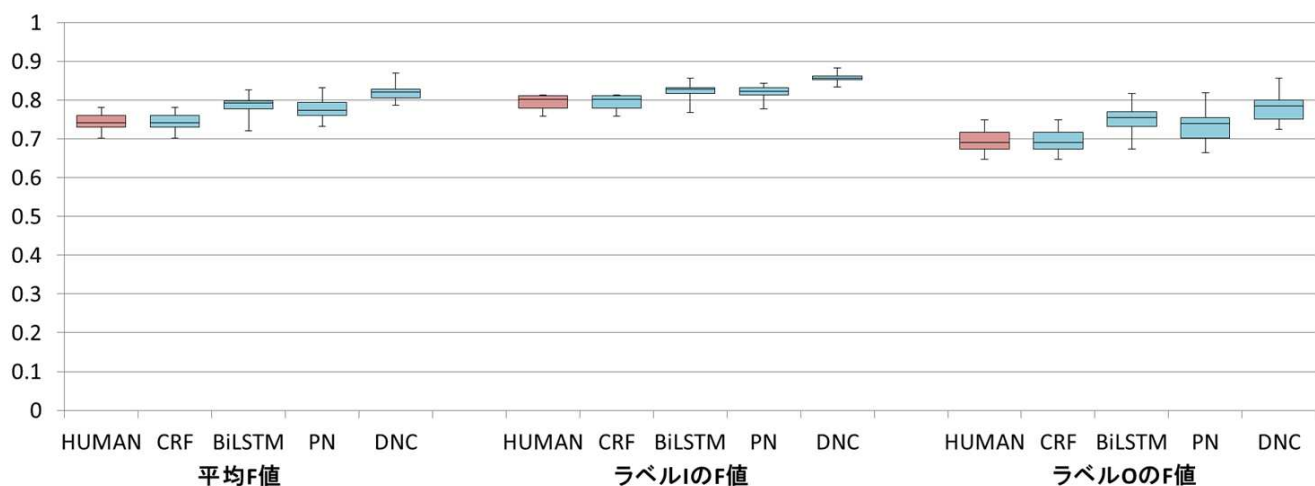


図 2.20: 訓練1人 - テスト4人（共通データのみ使用）

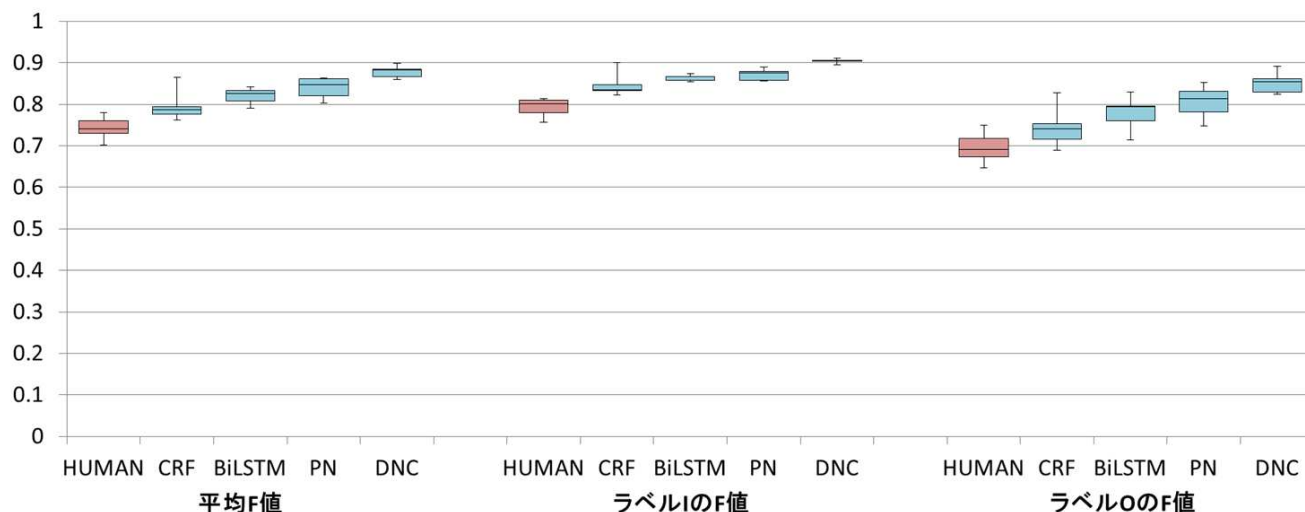


図 2.21: 訓練4人 - テスト1人 (全データ使用)

2.6.6 日経コーパスを用いた評価実験

2.6.6.1 実験設定

日経コーパスを用いて、提案モデルの性能を評価する。作業員 i のデータ 90 個を訓練、作業員 j のデータ 10 個をテスト ($i \neq j$) となるように分割し、10 分割交差検定で評価した。

事前学習なし、補助情報なしで単一作業員のデータを用いて学習したモデルを SCM1、これに事前学習を行ったモデルを SCM1_pre、さらに補助情報を加えて学習したモデルを SCM1+_pre、作業員 2 人のデータを用いて学習したモデルを SCM2+_pre とする。また、係り受け木の刈り込みによってあらかじめ妥当な圧縮文候補を生成しておき [Nomoto 08]、文圧縮モデルで推定した文節の含めるべき確率に基づき各圧縮文候補をランキングし、最上位の結果を採用したときの結果を SCM2+_pre_TDP とする。

比較手法として、2.6.4 節で説明した補助情報と文節を構成する単語の Bag-of-Words を素性として双方向 LSTM により文圧縮を行ったときの結果 (BiLSTM)、作業員の一方をシステム出力、他方を正解として Accuracy を算出した結果 (Human)、ランダムに文節を選択したときの結果 (Random)、文圧縮を行わなかったときの結果 (NoCompression) と比較を行った。

2.6.6.2 実験結果

Accuracy で評価した結果を図 2.22 に示す。この結果から提案手法が BiLSTM モデルよりも優れていることが分かった。また、文圧縮のタスクにおいても、事前学習なしのモデル (SCM1) よりも事前学習を行ったモデル (SCM1_pre) の方が良く、さらに補助情報を加えて学習するこ

とで (SCM1+_pre), 複数人のデータを活用することで (SCM2+_pre) 性能が向上することが分かった. 無圧縮の文の Accuracy (NoCompression) が高いのは, 作業者が作成した圧縮文の要約率が 70%程度であることに起因する.

係り受け木の刈り込みによって生成された候補のランキングに基づいて文圧縮を行った結果を図 2.23 に示す. この結果から, ランキングに基づく手法 (SCM2+_pre_TDP) の方が全体の Accuracy は向上するが, 文節を「省く」精度において系列ラベリングの結果 (SCM2+_pre) よりもわずかに劣ることが分かった. SCM2+_pre_TDP によって得られる圧縮文の質は係り受け解析の精度に影響されるが, 系列ラベリングで含めるべきとされた文節を単純につなげた文よりも文法的に正しいことが多く, 要約率の制御も可能であるため実用的であると考えられる.

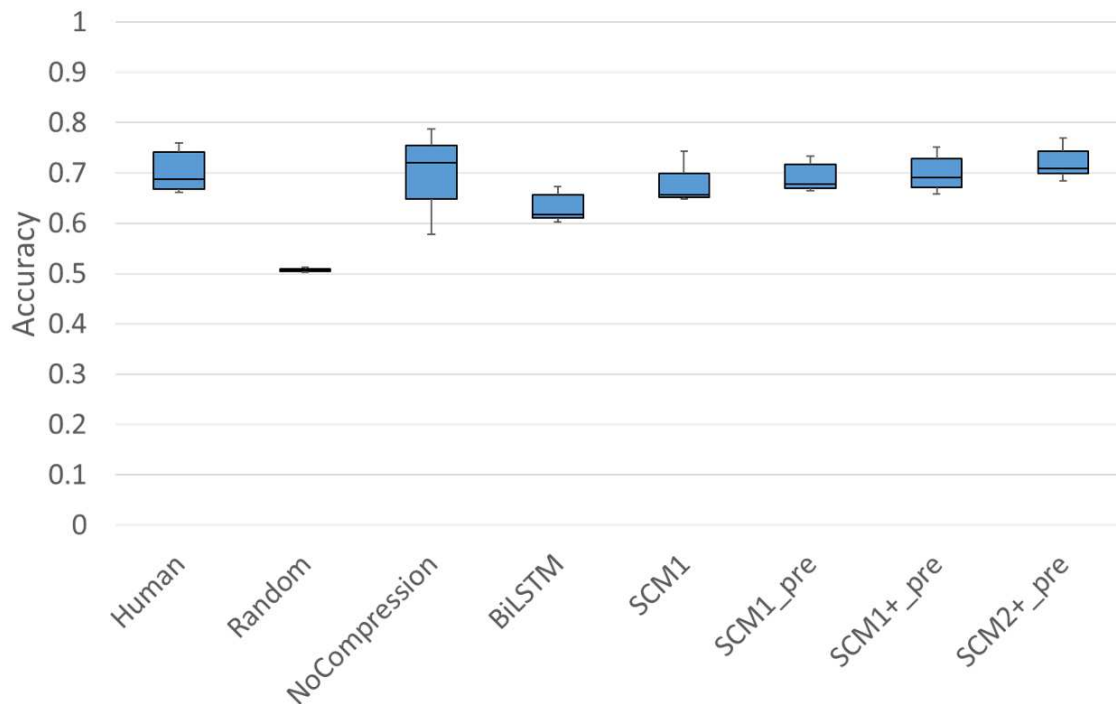


図 2.22: 文圧縮の実験結果

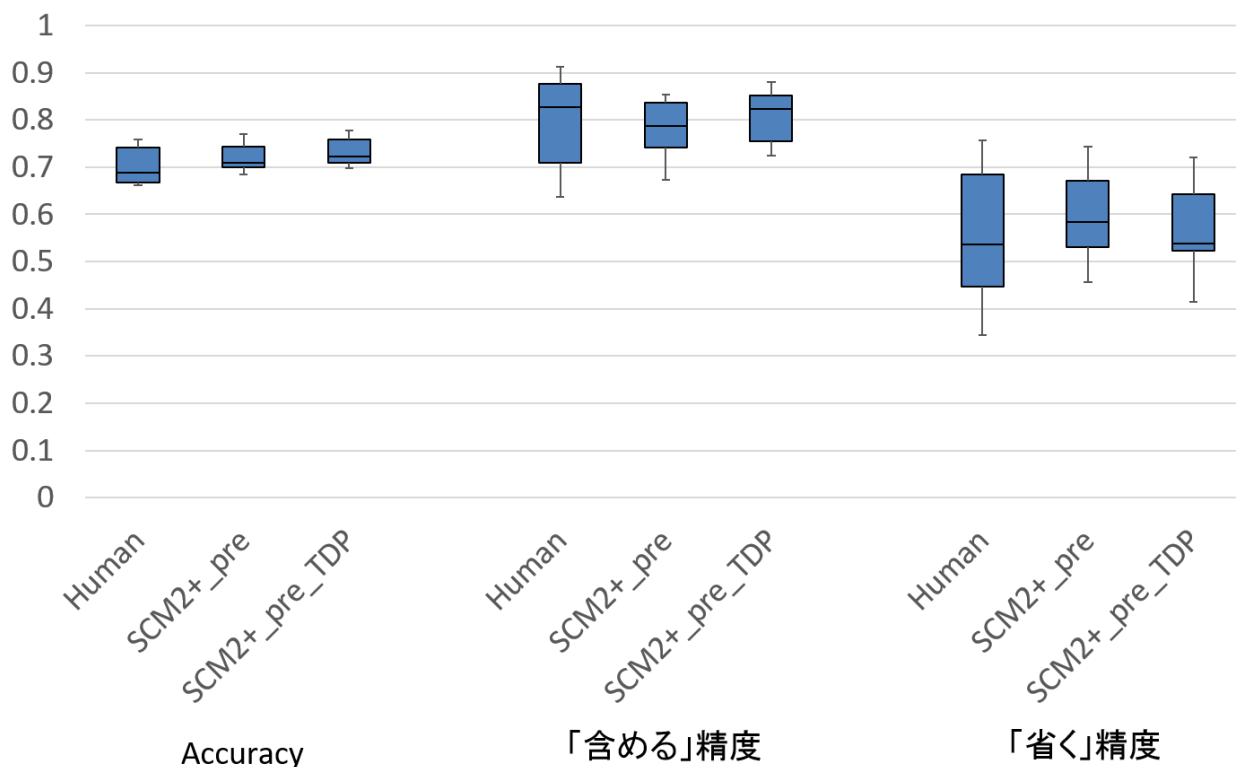


図 2.23: 文法性を保証した際の文圧縮の実験結果

2.6.7 用語の説明に関して

文圧縮で抽出が望まれる情報として、核となる情報の他に、知名度の低い用語に対する説明が挙げられる。例えば、図 2.24 のような原文から「Olli に Watson が搭載されたんだって」という発話が生成されたとすると、ユーザーは「Olli がバス」であるということを知っていなくてはならない。話し言葉の同時性により、ユーザーから「何それ」といった質問がくるかもしれないが、もし、ユーザーが質問せず、知らない用語をそのまま放置してしまったら、以降の話の理解に支障をきたす恐れがある。そのため、知名度の低い用語に対する説明は、なるべく文圧縮の段階で抽出される方が望ましい。逆に「Olli」という単語がニュースを理解するうえであまり重要でないとするならば、そもそも「Olli」という単語を出さずに「自動運転バス」と参照してしまってもよい。現状、その用語に対する説明が不要かどうかはコーパス作成時に作業者が判断している。しかしながら、用語の説明を加えるべきか否かについてはユーザーの知識レベルに応じて判断すべきであり、必ずしも文圧縮の問題に位置づける必要はないと考える。

EVスタートアップ Local Motorsが、 IBMの 人工知能 Watsonを 搭載する 自動運転バス 「Olli」を 発表しました。
S : あのさー、Olliって知ってる？
U : えっ、知らない
U : 何それ？
S : Olliっていう自動運転バスがあるんだけど
S : それにね、Watsonが搭載されるんだって
U : えっ、Watsonって、あれ？
U : えっとー、IBMの・・・
S : そうそう、IBMの人工知能

図 2.24: 会話利用を目的とした文圧縮の例 2

2.7 要約の品質評価

日経新聞のニュース記事に対して要約を行い、その結果について内容の一貫性や誤解を生じさせる表現の有無という観点で生成された要約の品質を評価した。評価で使用したニュース記事のカテゴリごとの統計を図 2.25 に示す。実験では抽出する重要文の数を 3 に設定し、重要文抽出アルゴリズムには MCKP を、文圧縮アルゴリズムには双方向 LSTM モデルを使用した。また、データセットとして Yahoo コーパスを使用した。

生成された要約の各文について非文がどれだけ含まれるか、また、誤解を生じさせる文がどれだけ含まれるか評価した結果を表 2.7、図 2.26 に示す。また、データ数が多い主要なカテゴリごとの評価結果を表 2.8、図 2.27 に示す。ここで、誤解を生じさせる文とは要約した結果、元の記事内容と異なる意味になってしまっている文を指す。誤解を生じさせる文の例として主語や条件節の欠落、述語のモダリティの誤り (e.g. 広まってきたのではないだろうか → 広まってこなかったんだって) などがある。なお、文圧縮は文節を抽出しただけで読みやすい表現に直す作業までは行っていないため、ここでは、口語化処理も行った結果に対して評価した。そのため、非文と評価された文には口語化処理の誤りに起因するものが多く見られた。非文として例えば、連用形の書き換え誤り (e.g. 「使い」 → 「使いて」 (正しくは「使って」)) や KNP の述語項構造解析誤りに起因する連体修飾節の展開誤り (e.g. 京都大学のボートを漕いだ青春の日々は → 日々は京都大学のボートを漕いだ青春で) などがある。

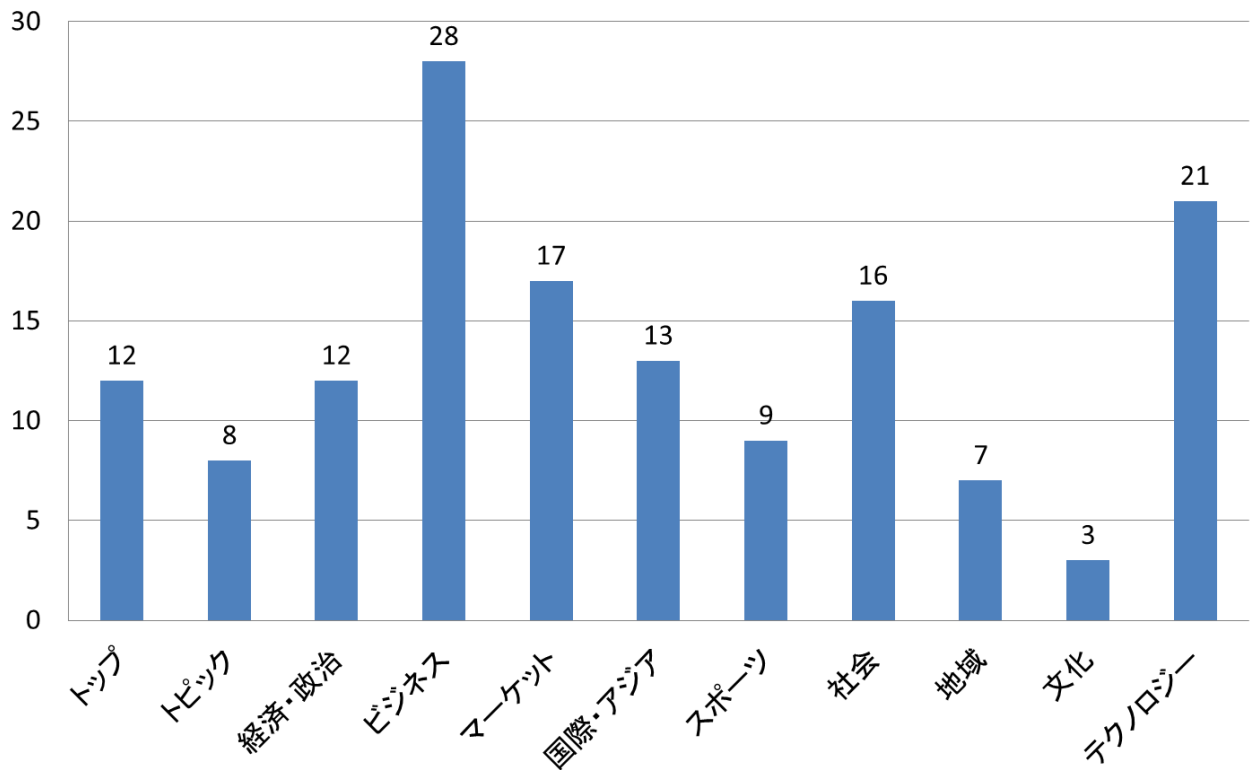


図 2.25: 要約の品質評価で用いたニュース記事の統計

表 2.7: 要約の各文に対する評価結果

非文	26
誤解を生じる文	13
正しい文	261

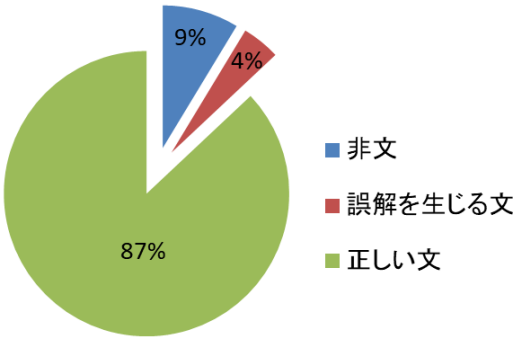


図 2.26: 要約の各文に対する評価結果

表 2.8: 要約の各文に対する評価結果（カテゴリごと）

	ビジネス	テクノロジー	マーケット	社会	経済・政治	スポーツ
非文	2	2	4	6	3	2
誤解を生じる文	4	4	3	1	1	0
正しい文	78	57	44	41	32	25

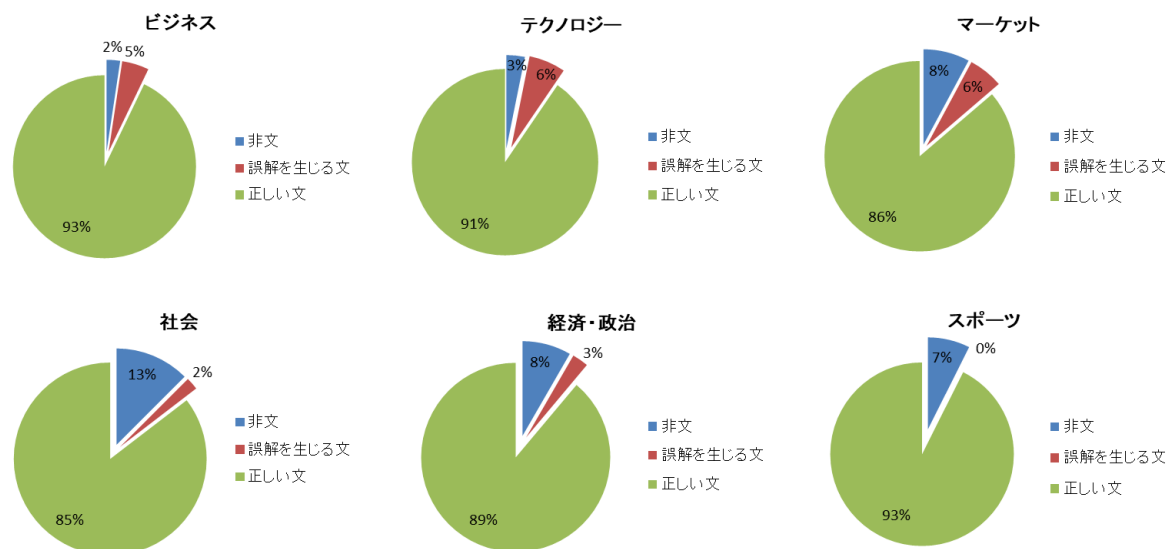


図 2.27: 要約の各文に対する評価結果（カテゴリごと）

次に、要約全体について内容が一貫性のない文章になっているか、誤解を生じさせる文章になっているか評価した結果を表 2.9、図 2.28 に示す。また、データ数が多い主要なカテゴリごとの評価結果を表 2.10、図 2.29 に示す。ここで一貫性のない文章とは、文章の途中で脈絡なく急に話題が変わっている文章を指す。一貫性のない要約は、元のニュース記事で複数の話題を扱っている場合に多く見られた。

表 2.9: 要約全体に対する評価結果

誤解を生じさせる文章	8
一貫性のない文章	6
正しい文	86

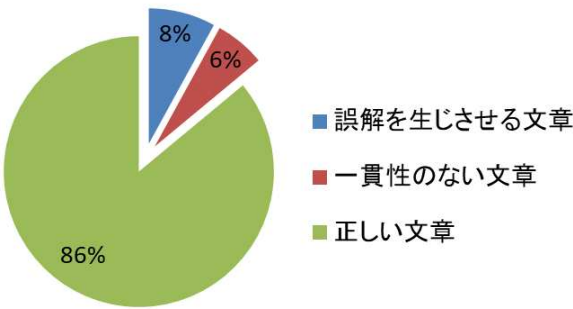


図 2.28: 要約全体に対する評価結果

表 2.10: 要約全体に対する評価結果（カテゴリごと）

	ビジネス	テクノロジー	マーケット	社会	経済・政治	スポーツ
非文	0	0	2	1	3	2
誤解を生じる文	2	1	1	0	0	1
正しい文	26	20	14	15	12	9

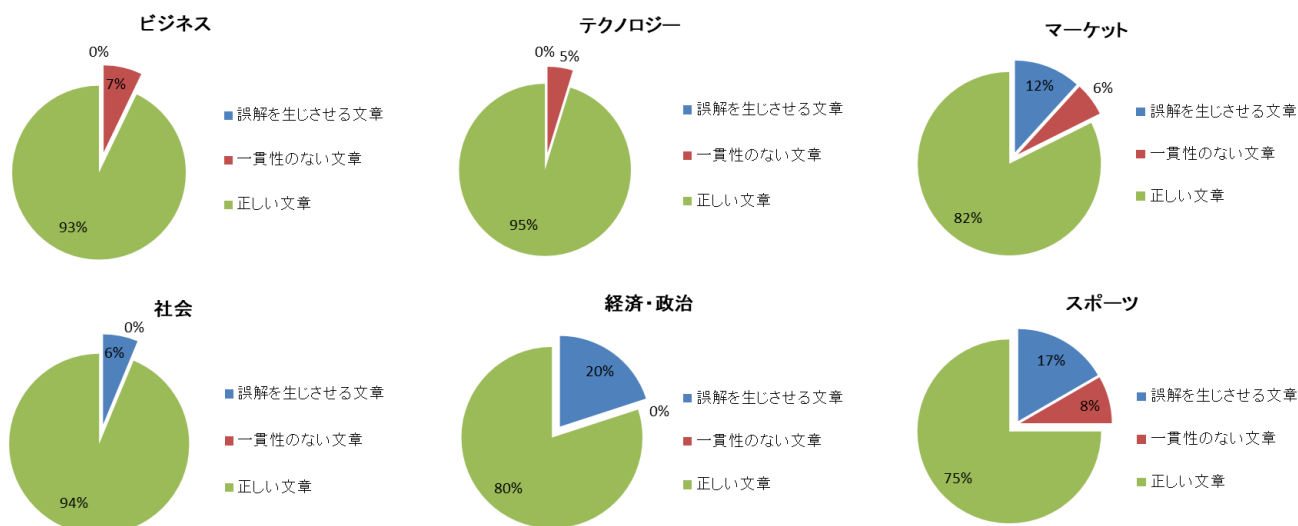


図 2.29: 要約全体に対する評価結果（カテゴリごと）

2.8 まとめと今後の課題

重要文抽出では、BERT を基本構造とした双方向モデルにより、文脈を考慮して単語の重要度を推定する手法を提案した。実験の結果、従来の文脈を用いない手法よりも、提案モデルで推定した単語重要度を用いたときの方が高い ROUGE-1 を示した。

整列では、複数記事からの要約に関して、ある記事の重要文と別記事の重要文前後の文章との類似度に基づいて前後関係を決める手法を提案し、有効性を確認した。

文圧縮では、文圧縮の問題を文節レベルの系列ラベリングの問題として定式化し、BERT に基づく文圧縮モデルを提案した。実験の結果、単純な双方向 LSTM モデルに比べて、提案モデルの方が高い精度を示した。また、係り受け構造を制約とすることで文法的に正しい要約文を生成できることを確認した。

今後は、ユーザーとの対話履歴を活用し、ユーザーごとに適切な要約を作成する方法について検討する。

第3章 会話によるニュース記事伝達のための口語化

3.1 序論

3.1.1 研究背景

聴覚メディアでニュース記事の内容を未加工のまま伝えと、聞き手はたとえ興味がない内容であっても最後まで黙って聴くしかなく、フラストレーションの大きいものとなる。そこで、要約した内容を会話的に質疑を交えながらインクリメンタルに伝える方法が考えられるが、視覚メディアによる情報伝達を前提として書かれたニュース記事の内容を会話で伝えるためには、書き言葉から話し言葉への変換が必要である。

書き言葉と話し言葉の最も基本的な違いとして、話し言葉の同時性がある [畠 87]。話し言葉によるコミュニケーションでは、話し手が情報を送信すると同時に聞き手が受信することになる。話す内容や話すペースは話し手が制御することになるため、聞き手の負担は、話し手の裁量に依存する。そのため、話し言葉では書き言葉に比べ、聞き手にとって負担の小さい表現が使われる傾向にある。

本章では、上述の話し言葉の特性を考慮し、書き言葉で書かれたニュース記事の内容を会話で伝えるのに適した表現に書き換える方法について検討する。

3.1.2 関連研究

平易化の従来研究については、付録 F.2 節にまとめた。

3.1.2.1 書き言葉から話し言葉への書き換え

書き言葉を話し言葉に書き換える研究は、主に音声合成の分野で行われてきた。音声合成を用いたアプリケーション開発では次の二つが問題となる。一つは、韻律やイントネーションが不自然であるという音響的な問題であり、もう一つは、音声化されるテキストの言語的な問題である。書き言葉から話し言葉への書き換えに関する研究は後者の問題に着目したものである。音声合成分野で話し言葉が扱われる場合、ほとんどの場合、聞き手を特定しない不特定多数の

人に（一方的に）情報を伝えることを想定しているため、敬体（ですます調）への書き換えが行われる。林らは、ニュース音声の自動配信アプリケーションを想定し、新聞記事を対象に文体変換規則による変換と体言止めの補完による話し言葉への変換を行った（e.g. 論争を展開 ⇒ 論争を展開しました）[林 07a, 林 07b, 林 08]。大泉らは、格解析結果を用いた名詞化された用言を含む表現の変換（機種依存文字 ⇒ 機種に依存する文字）と、国語辞典を利用した漢語調の平易表現への変換（e.g. 決定する ⇒ はっきり決める）、普通体及び文語調の口語調への変換（e.g. 仕方ない ⇒ 仕方ありません）を行った[大泉 03]。鍛冶らは、用言を対象として、国語事典から用言の言い換えペアを学習し、WWW から自動収集した書き言葉コーパス（新聞記事など）と話し言葉コーパス（チャットや掲示板など）を用いて、書き言葉語彙から話し言葉語彙への言い換え（e.g. 異様だ ⇒ 違う）を行った[鍛冶 04]。

3.1.2.2 「やさしい日本語」プロジェクト

外国人にもわかりやすい日本語「やさしい日本語」を作る研究が行われている[弘前大学, しまね国際センター, 全 13]。「やさしい日本語」は、「難しい言葉を避け、簡単な語を使う」「二重否定を避ける」「文末表現はなるべく統一する」などの規則に従い作成される。やさしい日本語でニュースを伝えるサイトとして News Web Easy [田中 13]¹がある。News Web Easy で配信されているニュースは、やさしい日本語ニュースへの書き換えの基準[田中 10]に従い、書き換え支援システム[美野 12]を用いて人手で作成されている。後藤らは、一般ニュースからやさしい日本語のニュースへの自動変換を目指し、分析を行った[後藤 14]。なお、5 万文をやさしい日本語（平易な日本語語彙）に書き換えた対訳コーパスとして SNOW が作成され[Maruyama 18, 山本 17]、公開されている²。

やさしい日本語を作るという点において話し言葉への書き換えと近いところがあるが、「やさしい日本語」は外国人にもやさしい読み物の作成を志向した研究であり、利用者とのカジュアルな会話（多少くだけた言い回しも許容される）の実現を目指す本研究とは異なる。

3.1.2.3 対話システムにおける発話生成

対話システム研究における発話生成は、事前に人手で用意した発話文を使用したり、Twitter 上のツイートをそのまま利用したりすることが多い[別所 12, 目黒 14, 高津 15]。ヒューリスティックな変形ルールでツイートの語尾の口調を修正するなどの書き換えは行われている[稲葉 14]が、文書のような書き言葉で書かれたまとまりある情報を会話で話すための表現に書き

¹<https://www3.nhk.or.jp/news/easy/>

²<http://www.jnlp.org/SNOW/T15>

換える研究はない。

最近では, Encoder-Decoder モデルを用いて End-to-End で発話生成を行う研究も多いが [Vinyals 15b, Shang 15, Sordoni 15a, Sordoni 15b, Serban 16], これらのモデルの学習には膨大な対話データが必要であること, 内容の一貫性が保証できないことから, 本研究では, ヒューリスティックな書き換えを行う。

3.1.3 想定するキャラクター

本研究では, 友達と会話するようなカジュアルな感覚で利用できるシステムを目指している。そのため, 敬体 (ですます調) ではなく, もっとくだけた表現に書き換える。システムの性別は特に規定しておらず, 中性的なキャラクターを想定している。そのため, 「～らしいわよ」 (女性的) や 「～だぜ」 (男性的) のような, 特定の性別・キャラクターを想起させるような表現への書き換えは行わない。中性的なキャラクターにしておくことで, 音声合成に対する制約を低減できる。仮に性別を男性と定めてしまうと女性声の音声合成器は使用できなくなってしまう。

3.2 会話表現への書き換え

3.2.1 口語化の前処理: 自然な文への復元

文圧縮の結果 (確率 0.5 以上の文節) を単純につなぎ合わせるだけだと非文になることが多い。ここでは, 文圧縮時に得られた文節の重要度 (系列ラベリングモデルの確率値) に基づいて, 重要な文節を定め, その重要文節を起点として必要な文節を補い, 自然な文への復元を行う。

~0.1	~0.2	~0.3	~0.4	~0.5	~0.6	~0.7	~0.8	~0.9	~1.0
------	------	------	------	------	------	------	------	------	------

(a) 文節重要度ランク

「Suica」など10種類のICカードと連携ICカードを「Newニンテンドー3DS」などの
NFCリーダーにかざすと乗車履歴を読み取り、距離などに応じてポイントが手に入る仕組み

(b) ランク8 (最大ランク9-1) 以上の文節 (下線の部分)

「Suica」など10種類のICカードと連携ICカードを「Newニンテンドー3DS」などの
NFCリーダーにかざすと乗車履歴を読み取り、距離などに応じてポイントが手に入る仕組み

(c) 制約のもと補完した後の文節 (「読み取り、」を追加)

図 3.1: 口語化の前処理の例

まず、文節を重要度に応じて、10段階のランクに分類する（図3.1の(a)）。そして、その文における最大ランク-1以上の文節を含めるように制約のもと文節を選択していく。例えば、図3.1の(b)のような文があったとき、最も重要度の高い文節のランクが9なので、そこから1を引いたランク8以上の文節を重要文節（下線付きの文節）とする。さらに、このままでは「乗車履歴を」で文が終わっていて不自然なので、係り先の述語「読み取り」を補う（図3.1の(c)）。文節を補う際の制約を以下に示す。

- 述語節が含まれているが、その必須格が省かれていた場合、これを含める（Wikipediaに対して係り受け解析を行い、最も最頻出で述語にかかっていた格を必須格としている）。
- 述語節が含まれており、その述語節が一つしか格を持たず、その格が省かれていた場合、これを含める。
- 名詞節の係り先の述語節が含まれていない場合、係り先の述語節を含める。
- ただし、名詞節は係り先が存在しなくても、その名詞をゼロ代名詞とする述語節が含まれていれば、係り先を補わなくてよい。
- 同格やノ格は係り先の格を継承して係り先を削除できる（e.g. 「New ニンテンドー 3DS」などの NFC リーダーに ⇒ 「New ニンテンドー 3DS」などに）。
- 「」『』内の文節が名詞句である場合、一部が選ばれたならば、全て含める。

また、再帰的に遡って係り元を含めた方が自然な場合が多いので、現在は係り元のノ格や連体修飾節を含めるようにしている。ただし、「。」や「述語+、」「述語+接続助詞」「連用形述語」に到達したら止めるようにしている。「連れ+去る」のような複合述語は切らない。

3.2.2 口調の書き換え

神尾は話し手と聞き手がある文の情報をどの程度知っているかによって日本語の文末形式が変わるとする「情報のなわ張り理論」を提唱した [神尾 90, 神尾 02]。1990年のモデルでは、聞き手にとってその情報がなわ張りの内か外か（既知か未知か）、話し手にとってその情報がなわ張りの内か外かで異なる文末形式が顕在化するとした [神尾 90]（表3.1）。直接形は断定・言い切り（e.g. 「昨日は動物園に行って来ました」）、間接形は伝聞（e.g. 「パリの冬は寒いらしい」）、直接ね形は直接形に終助詞「ね」を伴ったもの（e.g. 「いい天気ですねえ」）、間接ね形は間接形に終助詞「ね」を伴ったもの（e.g. 「君は退屈そうだね」）である。

表 3.1: 「情報のなわ張り理論」[神尾 90]における文末形式の分類

		話し手のなわ張り	
		内	外
聞き手の なわ張り	外	直接形	間接形
	内	直接ね形	間接ね形

システムがユーザーにニュースを伝える場合を考える．ニュースの内容はユーザーにとって概ね未知の情報であるため，なわ張りの外側の情報として受け取られると考えられる．一方，システムは，ニュース内容の真偽を直接確認したわけではないので，基本的にはなわ張りの外側として伝えることになる．しかしながら，情報を咀嚼し自分のものになっているとも言えるため，なわ張りの内側として伝えることもできる．以上のことから，システムがニュースの内容をユーザーに伝える際は，伝聞口調で話すのが良く，場合によっては断定口調も使用できる．

以上のことを踏まえ，本システムでは，基本的には伝聞口調で情報を伝えるが，定義説明やオープンドメイン質問応答の際は，なわ張りの内側に存在する知識に基づいて応答するものと考え，断定口調を用いる．伝聞口調は「らしいよ」「んだって」「んだってさ」「そうだよ」「らしい」「みたい」などの中からランダムに選択している．ただし，同じ口調が連続しないように調整している．

3.2.3 動詞・名詞・形容詞の書き換え

動詞や名詞，形容詞の書き換えは辞書ベースで行った．例えば，動詞の書き換え（e.g. 死去する ⇒ 亡くなる）には ALAGIN の『動詞含意関係データベース』³を使用しており，同義語や類義語，表記ゆれに関する書き換え（e.g. 米国 ⇒ アメリカ）には ALAGIN の『基本的意味関係の事例ベース』⁴や日本語 WordNet⁵を使用している．さらに，略称や愛称への書き換え（e.g. アメリカ航空宇宙局 ⇒ NASA）には Wikipedia から自動で抽出したデータを使用している．これらの辞書を用いて，書き換え候補を抽出し，Streaming API⁶で収集したツイートデータにおいて高頻出なものへ書き換える [高津 16]．

³<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-2>

⁴<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-9>

⁵<http://nlpwww.nict.go.jp/wn-ja/>

⁶<https://dev.twitter.com/streaming/overview>

3.2.4 連体修飾節の展開

あまり重要でない連体修飾節は文圧縮の段階で省かれていることが望ましいが、文圧縮の段階で省けなかった場合やその情報が重要であった場合、連体修飾節が残っていることがある。連体修飾節が長い文をそのまま読み上げてしまうと、聞き手は被連体修飾節が表れるまで、曖昧な情報を記憶し続けなくてはならないため、聞き手にとって負担が大きい情報伝達となってしまう。そこで、口語化処理の段階で、連体修飾関係を検出した際、KNP⁷の述語項構造解析結果に基づいて書き換えを行った。

- 連体修飾節の用言が被連体修飾名詞を格に持つ場合
 - － 被連体修飾節が原文において主語または主題である場合：e.g. 投手指名練習に参加した藤浪は ⇒ 藤浪は投手指名練習に参加したんだけど、
 - － 被連体修飾節が原文において主語・主題以外である場合：e.g. カナダ大会で連覇を狙うなでしこジャパンに ⇒ なでしこジャパンはカナダ大会で連覇を狙ってるんだけど、そのなでしこジャパンに
 - － ただし、連体修飾節が、被連体修飾名詞（固有名詞でない）の意味を限定するような働きをしている場合、展開しない：e.g. 人とのコミュニケーションを通じて成長する次世代オートバイ ⇏ 次世代オートバイが人とのコミュニケーションを通じて成長するんだけど、
- 体言止めの文の場合：e.g. 和製スーパーカーと言われた初代NSX ⇒ 初代NSXは和製スーパーカーと言われた
- 被連体修飾名詞がキーワードで外の関係にあり、その上位語が分かっている場合：e.g. 駅構内を撮影した写真を画像加工してスマートフォンなどに配信する「駅視－vision」を ⇒ 駅視－visionっていうサービスがあるんだけど、それって、駅構内を撮影した写真を画像加工してスマートフォンとかに配信するサービスなんだけど、そのサービスを

3.2.5 同格表現の書き換え

ニュース記事では、名詞が並ぶタイプの同格表現が多く用いられる。しかしながら、このような表現を会話でそのまま読み上げてしまうと、堅苦しい話し方になってしまう。そのため、話し言葉として適切な表現に書き換える必要がある。名詞が並ぶタイプの同格表現（名詞A＋名詞B）を会話で伝える場合、次の3つのケースが考えられる。

⁷<http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

- (1) 名詞 B の知名度が高い場合、名詞 A を省くことができる。(e.g. テニス選手 | 錦織圭が ⇒ 錦織圭が)
- (2) 名詞 B の知名度が低く、記事において重要でない場合、名詞 B を省くことができる。(e.g. タクシー運転手 | 山田一郎が ⇒ タクシー運転手が)
- (3) 名詞 B の知名度は低い、記事において重要な場合、「B っていう A」のように書き換える。(e.g. 自動運転バス | Olli を ⇒ Olli っていう自動運転バスを)
- 現在のシステムでは、知名度の判断は行っておらず、文圧縮の結果をそのまま採用している。なお、同格表現の判定には、KNP の係り受け解析結果を使用している。

- 後続表現が「ガ格」である場合

- 後続表現が「人名」である場合：e.g. 映画監督アーサー・ヒラーが ⇒ アーサー・ヒラーっていう映画監督がいるんだけど、その人がね
- 後続表現が「人名」以外である場合：e.g. フジテレビ系の番組「SMAP×SMAP」が ⇒ SMAP×SMAP っていうフジテレビ系の番組があるんだけど、それがね、

- 後続表現が「ガ格」以外である場合：e.g. 高級スポーツカー「NSX」を ⇒ NSX っていう高級スポーツカーを

- 連体修飾を伴っている場合

- 後続表現が「人名」である場合：e.g. 第 43 回アカデミー賞の監督賞にノミネートされた映画監督アーサー・ヒラーが ⇒ アーサー・ヒラーっていう映画監督がいるんだけど、その人、第 43 回アカデミー賞の監督賞にノミネートされた人なんだけど、その人がね、
- 後続表現が「人名」以外である場合：e.g. 日本郵便が駅などに設置している宅配ロッカー「はこぼす」を利用した… ⇒ はこぼすっていう宅配ロッカーがあるんだけど、それって、日本郵便が駅とかに設置してるロッカーなんだけど、それを利用した…

3.2.6 重文の発話分割

一回の発話が長いと聞き手の負担が大きくなってしまうため、発話はなるべく短い方が良い。ここでは、特に重文に着目し、特定の条件を満たすとき、複数の発話に分割する。本システムでは、特に「動詞連用形」、「動詞＋逆接の接続助詞、」に関して書き換えを行った。

- 後続表現が長い場合（例えば、格を 2 つ以上有するなど）：e.g. 大阪工業大学でドローン操縦者養成スクールが開校し、開校式でドローンのテスト飛行が公開された ⇒ 大阪工業

大学でドローン操縦者養成スクールが開校したんだって、それでね、開校式でドローンのテスト飛行が公開されたそうだよ

- 先行表現と後続表現の一方、または両方が短い場合、分割しない：e.g. 調査を継続し、「今回確認できなかった禁煙の効果についても見極めたい」としている ⇒ 調査を継続して、今回確認できなかった禁煙の効果についても見極めたいって
- 連体修飾節内の重文は分割しない：e.g. 飲食店やコンビニエンスストアなどに「認証型コンセント」の端末を設置し、スマートフォンやp cの充電を可能にする公衆電源サービス「e s p o t」の実証実験が8月23日に始まる ⇒ e s p o tっていう公衆電源サービスがあるんだけど、それって、飲食店とかコンビニエンスストアとかに認証型コンセントの端末を設置して、スマートフォンとかp cの充電を可能にするサービスなんだけど、そのサービスの実証実験が8月23日に始まるらしいよ
- 後続表現の述語が、先行表現に出現する主格以外の名詞を格に持つ（指示詞やゼロ代名詞も含む）場合、分割しない：店舗でプリペイドカードを購入し、（プリペイドカードを）専用コンセントにかざすと100円で20分間使用できる ⇒ 店舗でプリペイドカードを購入して、専用コンセントにかざすと、100円で20分間使えるそうだよ

3.2.7 能動態 ⇒ 受動態, 他動詞 ⇒ 自動詞

「を発表する」や「を発売する」、「を始める」や「を見つける」などの特定の述語表現に関して、原文で存在していた主語が文圧縮によって省かれた場合、「が発表される」や「が始まる」のように受動態や自動詞に書き換える。例えば、「Uberは9月28日、フードデリバリーサービス「UberEATS」を29日午前11時から都心部で始めると発表した。」という文において、文圧縮で主語が省かれなかった場合、「UberがUberEATSっていうフードデリバリーサービスを都心部で始めるらしいよ」のように口語化されるが、主語の「Uber」が省かれた場合、「UberEATSっていうフードデリバリーサービスが都心部で始まるらしいよ」となる。

3.2.8 名詞並列関係の書き換え

名詞Aと名詞Bが「A、B」のような並列関係にあるとき、以下のように書き換える。

- AとBがクラスとインスタンスの関係にある場合：AのB（e.g. ギタリスト、ブライアンメイさんが ⇒ ギタリストのブライアンメイさんが）
- AとBが例示である場合：AとかB（e.g. スマートフォン、PCなどの ⇒ スマートフォンとかPCとかの）

3.2.9 コメントの書き換え

コメントの内容は、文圧縮によって一部削られることがあるが、基本的に口語化処理でその内容を書き換えない。その代わり、コメント周辺の表現を書き換える。例えば、「大河原医師はこのゲームについて「…」と語った」という文があったとき、「大河原医師がこのゲームについて言ってただけど、…だって」のような表現に書き換える。

3.2.10 その他の表現の書き換え

- 基本連用形 ⇒ タ系連用テ形：e.g. 配置し ⇒ 配置して
- 進行形表現の書き換え：e.g. ～している ⇒ ～してる
- 逆接表現の書き換え：e.g. ～したが ⇒ ～したんだけど
- 副助詞の書き換え：「や」「など」⇒「とか」、「のみ」⇒「だけ」
- 接尾辞の書き換え：「ら」⇒「たち」
- 主題の名詞が未出であった場合ガ格に置き換える：「は」⇒「が」
- 述語の補完：e.g. ～という声も ⇒ ～という声も + ある、～かも ⇒ ～かも + しれない
- ノ格の挿入：e.g. 当初価格 ⇒ 当初の価格、交通系 IC カード ⇒ 交通系の IC カード
- 複合名詞の書き換え：e.g. 着用者 ⇒ 着用した人
- 「という」⇒「っていう」

3.2.11 述語の時制推定

「～を公開した。」のように言語表記から時制を推測できることもあるが（この場合は「過去」）、「～を公開し、」のような連用形や「～を公開。」のような体言止めの場合、言語表記だけで時制を推測するのは困難である。そこで、コーパスに対して「現在」「過去」「進行」の時制を付与し、機械学習の手法で推定する方法を取った。ここで、時制として、「現在」「過去」「進行」単体の他、「現在」と「進行」、「過去」と「進行」のような複数の時制タグが付与されることもある。

データ数は、「現在」が940個、「過去」が1200個、「進行」が180個である。識別関数として線形カーネルSVMを使用し、「現在」か「過去」かの二値分類問題と「進行」かどうかの二値分類問題の識別を行った。素性には、対象文節の形態素情報（表層形、品詞大分類、品詞細分類、活用形、活用型、ドメイン、カテゴリ）、係り先の文節が述語である場合その文節の形態素情報、文末文節の形態素情報、文中の時間表現の形態素情報を利用した。

10分割交差検定で評価した結果を表3.2、表3.3に示す。およそ90%程度のF値を示した。実

際に使用するときは、二つの識別器の結果の組み合わせで述語の時制を決定する。

表 3.2: 「現在」か「過去」か

	精度	再現率	F 値
現在	90.7	92.2	91.4
過去	93.9	92.6	93.2
平均	92.3	92.4	92.3

表 3.3: 「進行」かどうか

	精度	再現率	F 値
進行である	90.7	92.2	91.4
進行でない	93.9	92.6	93.2
平均	92.3	92.4	92.3

3.3 主観評価実験

主観評価にて、口語化結果に違和感がないことを確認する。

3.3.1 実験設定

13 人の被験者に対してアンケート評価を実施した。アンケートでは、作業者 1 の文圧縮結果に対して口語化処理を施した結果と作業者 1 が作成した口語化結果をそれぞれ 10 トピックずつ混ぜたものに対して評価させた。評価は以下の 3 つの項目から選択させた。

1. 違和感あり(許容できない)

- 言葉遣いが支離滅裂で理解できない。または、理解にかなりの労力を要する。

2. 違和感あり(許容できる)

- 若干言葉遣いに不自然な部分はあるが、意味は十分伝わる。

3. 違和感なし

- 自分でも同じような伝え方をする。

なお、アンケートでは、ニュース内容の難易度に左右されないように、特に言葉遣いについて評価するように強調した。実際に提示したアンケート内容を図 3.2 に示す。

【依頼内容】

ニュース内容を会話で話した際の発話文を計20つ用意しました。
各発話文の「言葉遣い」(注)に違和感があるか否かを評価してください。
※ニュース内容に専門用語などが使われる等して理解しがたい場合もありますが、
ここでは「内容」ではなく「言葉遣い」に関して評価してください。

(注):「言葉遣い」とは、物の言い方/言葉の使いぶりであり、例えば以下の赤字部分が該当します。

ubereats**っていう**フードデリバリーサービスが都心部で**始まるみたいだよ**
人気レストランの宅配メニューをスマートフォンから**注文できるんだって**
uberの従業員**じゃなくても**、審査に**通れば**、**配達して報酬がもらえるそうだよ**

【評価方法】

各トピックについて該当する評価を以下の3つの項目から選択し、
選択した番号を に記入してください。

1. **違和感あり(許容できない)**
言葉遣いが支離滅裂で理解できない。または、理解にかなりの労力を要する。
2. **違和感あり(許容できる)**
若干言葉遣いに不自然な部分はあるが、意味は十分伝わる。
3. **違和感なし**
自分でも同じような伝え方をする。

もし、違和感の理由が明確な場合はその理由を に記入してください。(記入は任意です)

以下、「言葉遣い」の評価をお願いします。計20トピック

トピック1 評価 → 違和感の理由(任意) →

印鑑はんこshopハンコズがニャン札プレートという猫のイラストが入った表札の発売を始めたんだって

かわいいプチ看板とかネコ部屋のドアプレートとしても使えるみたいだよ

かわいいネコの表情とかしぐさを集めた35種類からイラストを選べるらしい

無料オプションとして肉球マークを刻印できるんだってさ

図 3.2: アンケート用紙

3.3.2 実験結果

人間の口語化結果とシステムの口語化結果のそれぞれに対する各被験者の平均点の分布を箱ひげ図で表現したものを図3.3に示す。また、人間の口語化結果とシステムの口語化結果のそれぞれについて、全被験者の平均点に対し、平均を計算した値（総合平均点）を表3.4に示す。結果は、システムの方がわずかに良かった。これは、システムは汎用的な言い回しを目指して実装しているが、作業者による口語化結果は「こう話したい」というその人の個性に基づいて作られたものであり、被験者の個性との差異が大きかったためだと考えられる。実際、人間の口語化結果には「2020年の東京五輪を見据え、日本に来る外国人の利便性向上を図るんだね」のような少し癖のある表現も含まれている。

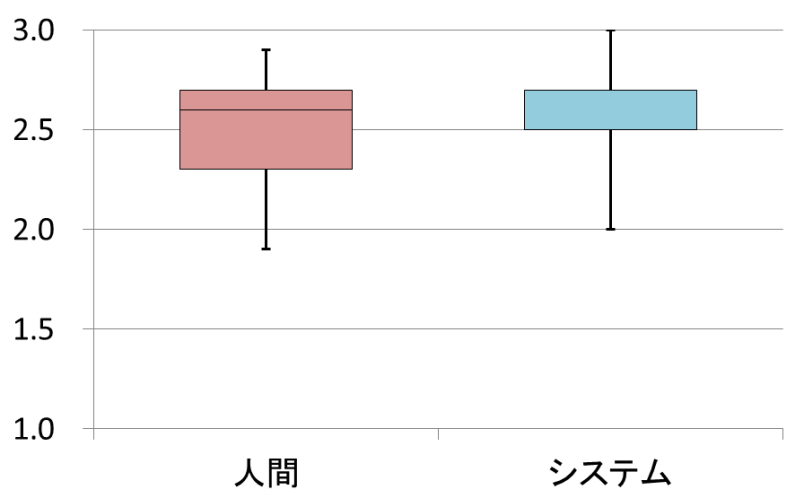


図 3.3: 各被験者の平均点のばらつき

表 3.4: 人間とシステムの総合平均点

	人間	システム
平均	2.50	2.62

3.3.3 違和感の理由

アンケートで「違和感あり」と回答した人に対して、違和感の理由が明確である場合は、その理由を任意で書くように指示した。違和感の理由として以下のような理由が挙げられた。

- 同じ口調が連続している
- 一人称と三人称が混在している（断定口調と伝聞口調の混在が原因）
- 発話のつなぎに接続詞がない
- 次に話したいことがあるのに発話権を維持しようとしていない（全ての発話が言い切り型になっている）

この他、「～だってさ」という言い回しは話し手自身、ニュース内容に興味が無いように感じると言ったコメントや途中でキャラクターが変わっているように感じたというコメントが見受けられた。さらに、「らしいよ」なら違和感はないが「らしい」には違和感があるという人や時制に違和感が覚えたという人もいた。

3.4 まとめと今後の課題

ニュース記事の文章は書き言葉で書かれており、この内容を会話で伝えるには、話し言葉に変換する必要がある。特に、会話による情報伝達では、送受信の同時生に起因する聞き手の負担を考え、やさしい表現での情報伝達が求められる。そこで、本章では、会話として違和感のない表現を作るべく、口調やその他表現の書き換え方法について検討した。そして、アンケートによる主観評価実験にて生成された発話の言葉遣いに違和感がないかどうかを確認した。結果は、人間が作成した口語表現に対する評価と同程度の評価値を得た。しかしながら、システムの口語化は作業者の文圧縮結果に対して行ったものであり、いわば文圧縮の精度が100%の元での評価結果である。実際は文圧縮した結果に対して行われるので、文圧縮による誤りがそのまま伝搬される。そこで、文圧縮時の誤りの影響を緩和すべく口語化を行う前に自然な文への復元処理を行っているが、その処理によって冗長な表現が作られてしまうという問題があり、いかに最小限の補完で非文を回避するかが今後の課題である。

本研究では、友達のようなカジュアルな関係で情報を伝えることを意識して、口語化のルールを作成した。しかしながら、7章のシステム評価時に行ったアンケート調査では馴れ馴れしいという意見もあった。そのため、今後は、丁寧な口調も導入し、ユーザーとの親密度に応じて、徐々にくだけた口調に変化させていくような制御方法について検討する。

個別の観点での課題を以下にまとめる。

前処理

口語化の前処理として、文圧縮後の文を自然な文へ復元する処理を行うことで、非文になるリスクは低減できるが、場合によってはこの処理により、冗長な表現が作られることがある。せっかく文圧縮で情報を削ったのに、これでは本末転倒であるため、非文にならず冗長にならない最低限の補完にとどめるように工夫する。

口調とモダリティ

ニュース記事では客観的な事実が中心に記述されているため、概ね伝聞口調で問題ないが、ブログ記事や評論文などでは、筆者の主観的な意見や推測が中心であり、それらを伝聞口調だけで伝えるのは不自然である。そのため、今後は、文章中の推量や疑問のような文のモダリティをどのように扱うかについて検討する必要がある。

連体修飾節の展開

連体修飾節の展開は、KNPの述語項構造解析結果の精度に大きく依存する。例えば、本来外の関係であるのに、述語の格と推定されてしまった結果、「高温の銅をビッグマックにかける動画も ⇒ 動画が高温の銅をビッグマックにかける」（被連体修飾語の「動画」が連体修飾語「かける」の主格と推定された）のように、誤った意味の文章が生成されてしまうことがある。一方、誤りではないが、連体修飾節を必ず展開するようにしてしまうと「ビッグマックの不思議な魅力に取りつかれる人が ⇒ 人がビッグマックの不思議な魅力に取りつかれる」のような過剰な展開が発生してしまう。そのため、今後は、コーパスを作成し、機械学習の手法で連体修飾節を展開するかどうかを判定する方法について検討する。

重文の発話分割

例えば、「世界初の手術ロボットによる目の内部の手術が英国で行われ、網膜疾患で右目の中央の視野が欠けていた70歳の患者が視力を取り戻した」という文があったとき、「行われ、」という連用形の述語節で発話を区切るかどうかという問題がある。発話を区切らずに一回の発話で話す場合、「ロボットを使った目の手術が世界で初めて行われたんだけど、70歳の患者が視力を取り戻したらしいよ」となり、二つの発話に分けて話す場合、「ロボットを使った目の手術が世界で初めて行われたんだって」「でね、70歳の患者が視力を取り戻したって」となる。重文を複数の発話に区切るかどうかをルールベースで判定するのは難しいため、コーパスを作成し、機械学習の手法で識別する方法を検討する。

固有表現の略称獲得

一般名詞の書き換え（e.g. 昨年 ⇒ 去年）は、適宜辞書に追加していけば、いずれ収束すると考えられるが、固有名詞は日々増え続けているため、常に辞書を更新し続けなければならない。人手で毎回更新するのはコストが大きいため、Wikipediaなどの言語資源を利用して略称を獲得する方法について検討する。

国名略称の書き換え

「トランプ米大統領」「米アップル」「米グーグル」を会話でそのまま読み上げるのは不自然である。これらは「米」を省略できるが、「米株式市場」「米国防総省」「米史上」の「米」は省略できない。省略できるかどうかの一つの判断基準として人名や組織名の固有名詞に接続しているかどうかを考えられる。しかしながら、固有名詞の知名度が低い場合はどこの国の企業か分からなくなってしまうので、用語の知名度も考慮すべきである。また、略された国名は音声的に伝わりにくい問題もある。その場合は略称を通称や正式名称に変換すべきである（e.g. イコスタ・クルーズ ⇒ イタリアのコスタ・クルーズ）。ただし、略称を通称に置き換えられる場合（e.g. 米政府 ⇒ アメリカ政府）と置き換えられない場合（e.g. 日米和親条約）があるので、これらを区別する仕組みも必要である。

ノ格の自動補完

「当初価格 ⇒ 当初の価格」のようにノ格を補った方が自然な場合がある。どのような場合にノ格を挿入するのが自然かを判別する必要がある。最も単純な方法としては、「当初価格」の頻度と「当初の価格」の頻度を比較し、頻度が高い方を採用する方法が考えられる。

述語の含意表現への変換

「死去する」というような書き言葉特有の表現を会話で使っても不自然でない表現（e.g. 「亡くなる」）に書き換える。この問題についても一般名詞と同様に、辞書に適宜追記していくことでいずれ収束すると考えられるが、論文 [高津 16] で検討したように ALAGIN の『動詞含意関係データベース』を用いた含意表現への書き換え手法も、今後取り入れたいと考えている。

第4章 副計画生成

4.1 序論

快適な情報享受を実現するためには、ユーザーの情報要求に対して即座に応じられることが重要である。もし仮に、質問を受けた時点で、回答の探索や会話表現への加工、音声合成を行い、その処理に時間がかかってしまうと、会話のリズムは崩れ、会話インタラクションは、フラストレーションが大きな作業になってしまう。そこで、迅速な応答を保証する枠組みとして、副計画を設計する。

副計画は主計画の情報を補うための発話計画である。副計画では、予想される質問に対する回答をあらかじめ用意しておき、ユーザーの情報要求に応じて、副計画の内容を提示する。一般に、質問のタイプは、大きくファクトイド型質問とノンファクトイド型質問に分類される [奥村 13]。ファクトイド型質問は、単純な事実や出来事などについて一言で答えられる質問である。一方、ノンファクトイド型質問は、人物や物事の定義を尋ねる質問（定義型）、理由や原因に関する質問（why 型）、方法や手続きに関する質問（how 型）、意見や評価を尋ねる質問など、一言で答えられない質問である。本研究では、ファクトイド型質問としては、真偽を尋ねる質問（真偽判定型）と固有表現に関する質問を扱う。また、ノンファクトイド型質問としては、定義型と why 型を扱う。この他、ユーザーの興味に応じた情報提供を行うために、類似度に基づく補足説明の機能を導入する。

以下、質問応答および補足説明の機能について説明する。

4.2 副計画の生成方法

4.2.1 質問応答

4.2.1.1 固有表現に関する質問応答

KNP の固有表現クラスや JUMAN の品詞細分類や名詞カテゴリ、何 + 助数詞（e.g. 何歳）などの情報に基づいて回答する。

「いつ」や「誰に」のような述語が非明示的な質問に対しては直近の述語項構造に基づいて回答する。「いつ行われたの」のような述語が明確な質問に対しては、直近の述語項構造で答えら

れない場合、記事全体から回答を探し、発話文と最も近くの候補を回答とする。さらに、「メキシコオープンはいつ行われたの」のような明示性の高い質問（非疑問詞の格を一つ以上有するなど）に対しては、記事内からも回答が見つからない場合、オープンドメイン質問応答機能へ移行し、Wikipedia や他のニュース記事から回答を探す。オープンドメイン質問応答では、質問解析によって得られた質問タイプ（固有表現クラス）で回答候補を絞り込んだうえで、検索結果と質問文の類似度と回答候補（固有表現）の出現文脈に基づいてランキングしている。

さらに、文脈による回答可能箇所を増やすために、述語項構造の含意認識を行い、含意関係にある述語項構造間で、互いに不足する情報を補う。例えば、「メキシコオープンがアカプルコで行われた」という文と「メキシコオープンが2月28日に行われた」という文があるとき、前者に日付情報「2月28日に」を付与し、後者に場所情報「アカプルコで」を付与する。

4.2.1.2 why 型質問応答

理由を尋ねる質問に回答する。「ため」や「ので」などの特定のパターンに基づいて理由を獲得する。

「なんで」や「どうして」のような述語が非明示的な質問に対しては、直近の述語項構造に基づいて回答する。「なんで欠場したの」のような述語が明確な質問に対しては、直近の述語項構造で答えられない場合、記事全体から回答を探し、発話文と最も近くの候補を回答とする。さらに、「なんで羽生結弦は国際大会を欠場するの」のような明示性の高い質問（格を一つ以上有するなど）に対しては、記事内からも回答が見つからない場合、オープンドメイン質問応答機能へ移行し、Wikipedia や他のニュース記事から回答を探す。オープンドメイン質問応答では、原因と結果のペアからなるデータベースの結果部に対して検索をかけ、その原因部を回答候補とし、候補内での頻出単語を多く含むものが上位にくるようにランキングしている。

さらに、文脈による回答可能箇所を増やすために、理由表現を有する述語の述語項構造を用いて、含意関係にある他の文の述語項構造に対する理由の付与を行っている。例えば、「羽生結弦は腰の痛みのため、国際大会を欠場する。」という文から羽生結弦が国際大会を欠場する理由として「腰の痛みのため」という表現が獲得できる。別の記事に「羽生結弦はフィンランドで出場を予定していた国際大会を欠場する。」という文があったとき、述語項構造の含意関係を確認し、含意していれば、後者に「腰の痛みのため」という理由を与える。

4.2.1.3 定義型質問応答

用語の定義を問う質問に回答する．記事内からは，同格（e.g. IBM の人工知能「Watson」）や連体修飾（e.g. ソチオリンピックで金メダルを獲得した羽生結弦），直接的な記述（e.g. 「NP とは～」）を利用して定義を獲得する．連体修飾からの定義獲得では，被修飾名詞の固有表現（e.g. 羽生結弦 ⇒ 人物）または Wikipedia における上位語情報（e.g. 羽生結弦 ⇒ フィギュアスケート選手）を利用して，最近類と種差の考え方にに基づき，定義文を作成する（e.g. 羽生結弦は、ソチオリンピックで金メダルを獲得したフィギュアスケート選手だよ）．

記事外の定義情報として，Wikipedia の定義，EDR 辞書¹の定義，Web コーパス²から獲得した定義（C.5.3 節参照）を使用した．複数の定義文が候補となる場合もあるが，その場合は，定義候補集合内で出現する単語の頻度を計算し，頻出単語を多く含む定義文がランキング上位にくるようにしている．

「羽生結弦って誰」のような質問であれば，何について知りたいかが明確であるが，実際の会話では，「何それ」や「誰それ」といった形で質問されることが多い（7.1.6）．このような質問の場合，まず，指示対象を特定する必要がある．「何それ」や「誰それ」といった質問は，知名度の低い用語が出現したときに多く，知名度の低さは出現頻度の低さと関係が深い．そこで，これまでに発話した内容のうち，現在の発話と一つ前の発話に含まれる単語を候補として，Wikipedia や過去のニュース記事において最も出現頻度が低い用語について説明する．さらに，「誰それ」であれば人名に対する質問であることが分かるため，固有表現が PERSON の名詞に絞り込むことができる．一方で，「何それ」の候補は多いが，人名や日付を「何それ」と聞くことは少ないため，固有表現情報を用いてこれらを候補から除く．また，ニュース記事ではしばしばキーワードとなる用語を「」でくくっていることがあり，それに対して「何それ」と質問されることが多いことから，そのようなキーワードがあれば，その用語についての説明を優先する．

4.2.1.4 真偽判定型質問応答

ユーザーの発話内容が記事の内容と一致しているかどうかを判定する．述語項構造の含意矛盾認識により実現する．例えば「錦織圭がダビド・フェレールに敗れた」という文があるとき，「錦織（ガ）負けたんだ」というユーザー発話に対しては「そうだよ，錦織が負けたんだよ」などの肯定表現で応じ，「錦織（ガ）勝ったの」というユーザー発話に対しては「違う違う．錦織が負けたんだよ」のように，間違いを指摘したうえで，正しい情報を与える．また，格交替によって起こる矛盾（e.g. 「ダビド・フェレールが負けたの」）に対しても訂正を行う（e.g. 「い

¹http://www2.nict.go.jp/ipp/EDR/JPN/J_indexTop.html

²<http://s-yata.jp/corpus/nwc2010/>

や、ダビド・フェレールが勝ったんだよ」). 以下, 述語項構造間の含意矛盾認識の方法について説明する.

述語項構造の含意矛盾認識

述語項構造の含意矛盾認識は, 真偽判定やファクトイド型質問応答, why 型質問応答で必要となる技術である. ここで, 述語項構造が含意しているとは, 次の二つの条件を満たすことである.

[含意条件 1] 一方の述語が他方の述語を含意している

一方の述語が他方の述語を含意しており, 両方の述語が同じ極性 (肯定/否定) である (e.g. 敗北する → 負ける). または, 述語が反義関係にあり, 一方が否定表現である (e.g. 負ける → 勝てない). このとき, 述語が含意関係にあると考える.

[含意条件 2] 述語の格構造が類似している

全ての格に関して共通格の要素間で類似度を計算し, その平均が閾値以上であれば同一の述語項構造であると判断する. 閾値は格要素の数が少ないほど高くなるように設定する (例えば, 格要素が一つのとき 0.8, それ以上のとき 0.7 など). 格要素の表層的な類似度は正規化したレーベンシュタイン距離 (レーベンシュタイン距離を長い方の文字列の文字数で割った値) により計算する. ただし, 辞書上で同義関係にある単語 (e.g. 「米国」と「アメリカ」, 「原発」と「原子力発電」) は類似度 1 とする.

述語項構造が矛盾しているとは, 次の二つの条件を満たすことである.

[矛盾条件 1] 述語が反義関係にある

述語が反義関係にあり, 両方の述語が同じ極性 (肯定/否定) である (e.g. 負ける ⊥ 勝つ). または, 一方の述語が他方の述語を含意しており, 一方が否定表現である (e.g. 敗北する ⊥ 負けない). このとき, 述語が反義関係にあると考える.

[矛盾条件 2] 述語の格構造が類似している

含意条件 2 と同じである.

ここで, 動詞の含意知識として ALAGIN³の『動詞含意関係データベース』⁴を, 格要素の同義知識として ALAGIN の『基本的意味関係の事例ベース』⁵の異表記情報と同類語情報, 及び, NICT の『日本語 WordNet 同義対データベース』⁶を使用している. 反義知識は Web 上から人手で収集した反義ペアを『動詞含意関係データベース』を用いて拡張したものを使用している.

³<http://www.alagin.jp/>

⁴<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-2>

⁵<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-9>

⁶<http://compiling.hss.ntu.edu.sg/wnja/jpn/downloads.html>

4.2.2 補足説明

4.2.2.1 類似文の提示

ユーザー発話における内容語を一定以上包含する文（例えば3割）がある場合、その内容を伝える。ただし、優先順位は記事外質問応答よりは高く、記事内質問応答よりは低い。つまり、ユーザーの質問に対して記事内から回答が得られなかった場合でも、類似する文があれば、それを提示する。さらに、次の文が段落を超えていない場合、その内容も提示する。これは、単にユーザーが興味を示した情報を提示するだけでなく、関連する情報も提示することで、会話に広がりを持たせるためである。

4.2.2.2 類似コメントの提示

ユーザーの感想と類似するコメントがある場合、それを提示する。例えば、「でも、食べ物を無駄にしちゃダメでしょ」という感想に対して「実際、食べ物を無駄にするなって言ってる人もいる」のように類似コメントを提示する。さらに、別のコメントがある場合は、それも提示する（e.g. 「自家製のハンバーガーなどでも同じ実験をしてくれっていう声もあるみたいだよ」）。

4.2.2.3 情報の付加

ユーザーのトピックに対する興味が高そうな場合、システムが主体的に情報を付加する。現在のシステムでは、主に「面白いね」などの感想発話を受け付けた際に、興味ありと判定し、情報を付加するアルゴリズムになっている。ここで、システムの発話中または、システムの発話が終了したタイミングで、興味ありと判定された場合、次の文へ遷移する。

4.2.3 その他

4.2.3.1 名詞反復確認

システム発話中に含まれる名詞をユーザーが反復して発話した場合、肯定する。例えば、「ダイコンの辛みとかたくあんの匂いって、グルコラファサチンの分解産物によってもたらされるらしいよ」というシステム発話の後に、ユーザーが「グルコラファサチン」と発話したとき、「そう、グルコラファサチン」と肯定する。ただし、この計画へ遷移する優先順位は低く、当該名詞を主題として持つ文があれば、その内容を伝える「グルコラファサチンは…」（補足説明）。さらに、主題として持つ文がなくても、当該名詞がニュースにおけるキーワードであり、説明可能であれば説明する（定義型質問応答）「グルコラファサチンっていうのは、…」。また、用語の読みを一部包含する発話をした場合（e.g. 「グルコラなんだって」）、その内容を補間する。

(e.g. 「グルコラファサチン」).

4.2.3.2 復帰点

ユーザーはシステムが話している途中でも、割り込んで質問することができる。その際、システムは文脈情報を用いて上記アルゴリズムで回答するが、その後、言いかけた発話を言い直す。この際、最初から言い直すと、長い発話の場合、冗長になってしまうので、2 格前から言い直すようにしている。例えば、「製品とかサービスの受注額が目標を 11 % も上回る成果を出したらしいよ」という発話において、「上回る」の発話中に割り込みがあった場合、「目標を」から再度発話する。

4.2.3.3 同じような発話の回避

現在のシステムでは、一度提示した文は二度提示しないようにしている。例えば、まだ話していない主計画の内容を、ユーザーの質問によって先に提示した場合、その内容を主計画から取り除く。副計画に関しても同様である。ただし、質問応答によって文の一部だけが提示されたに過ぎない場合は保持する。また、異なる文であっても同じような内容の文（コサイン類似度が高い文、または、一方が他方を含意している文）がある場合、提示された文と類似する文を全て計画から取り除く。この仕組みは、特に複数記事から発話計画を作成するとき重要となる。

4.3 会話システムの応答分析

本章では、「何それ」に特化したシステムであるということを被験者に伝えたくてシステムと会話させ、そのときの質問応答の結果について分析を行った。

4.3.1 実験設定

テクノロジー系のニュース 10 記事について被験者 10 名（20～30 代の男性 8 名、女性 2 名）にシステムと会話させた。被験者には、質問応答として特に「何それ」に特化したシステムであることを伝えた。また、「えっとー」や「ちょっと待って」といった待機要求、「もう一回」や「なんて言ったの」といった反復要求などの音声コマンドの存在を伝えた。主計画を構成する重要文の数は 3 に設定した。

4.3.2 実験結果

音声認識で得られた372発話について分析を行った。これらの発話を「質問（何それ？）」「質問（〇〇って何？）」「質問（その他）」「音声コマンド」「相槌・感想・その他」「音声認識誤り」の6つに分類したときの各発話タイプの割合を図4.1に示す。このうち、音声認識誤り発話に対してシステムがどのような応答を行ったか以下の5つの項目に分類し、各割合を算出した結果を図4.2に示す。

- 正答（エラーハンドリング）

付録A章で説明するエラーハンドリングの機構によって正解した場合これに分類する。

- 回答不可

回答できず「分からない」と答えた場合これに分類する。

- 誤答

間違った回答した場合これに分類する。

- 無視（質問・音声コマンド）

音声コマンドや質問を無視した場合これに分類する。

- 無視（相槌・感想・その他）

相槌や感想、その他の発話を無視した場合これに分類する。

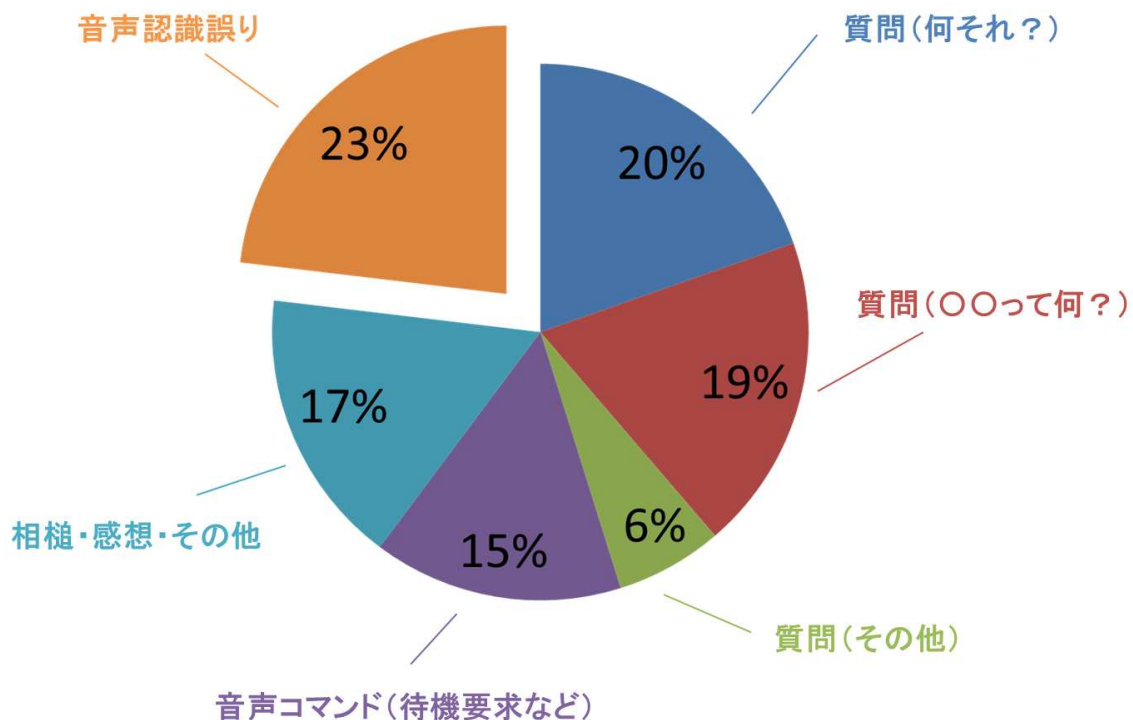


図 4.1: ユーザー発話の分類と各発話タイプの割合

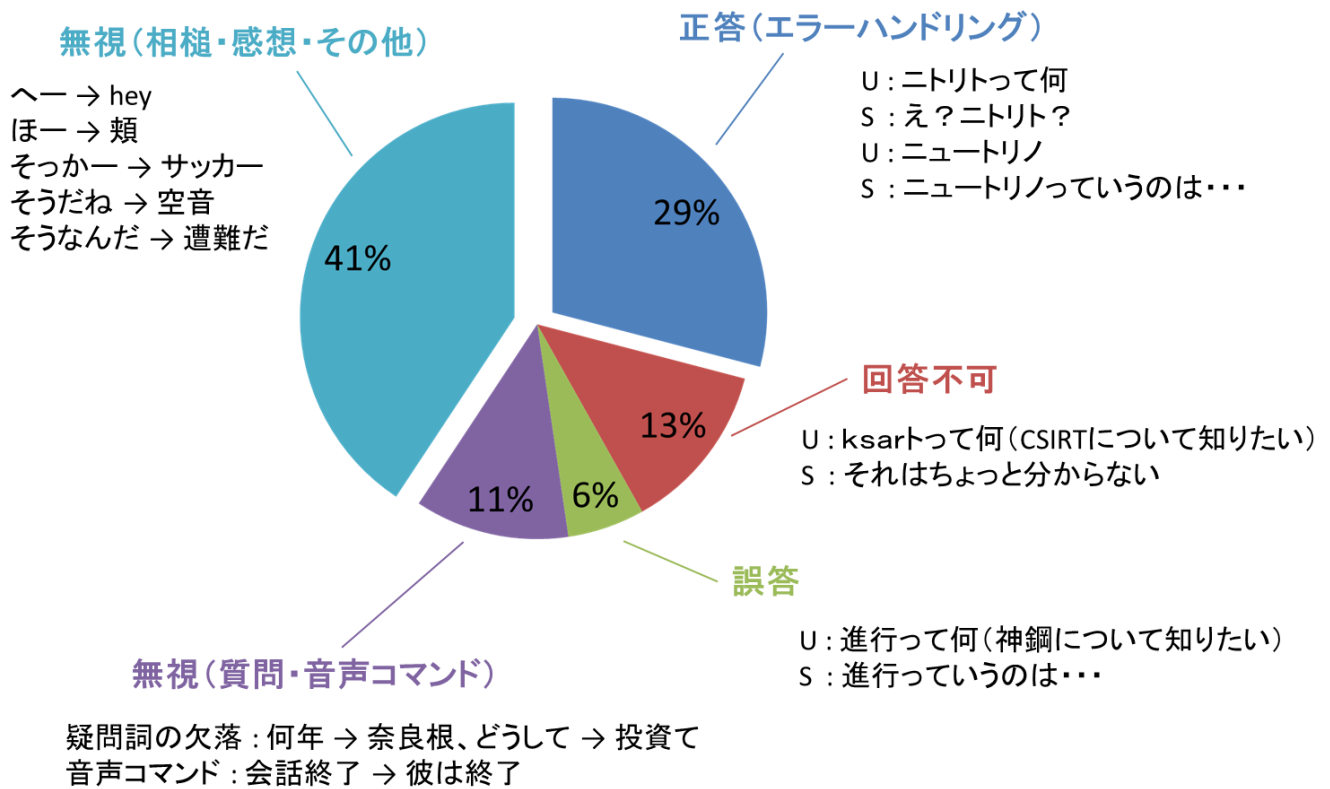


図 4.2: 音声認識誤り発話に対するシステムの応答分類

次に、「何それ？」の質問に対してシステムがどのような応答を行ったか以下の8つの項目に分類し、各割合を算出した結果を図 4.3 に示す。

- 正答

正しく回答できた場合これに分類する。

- 誤答(別の単語について聞いたかった。DBにない)

質問対象の用語がデータベースになく別の用語について回答してしまった場合これに分類する。

- 回答不可(未サポートの質問方法)

サポート外の質問方法のため回答できなかった場合これに分類する。

- 誤答(記事中で意味を知りたい)

記事中で意味を知りたかったが、一般的な定義を述べてしまった場合これに分類する。

- 誤答(別の単語について聞いたかった。DBにある)

質問対象の用語がデータベースにあるが、別の用語について回答してしまった場合これに分類する。低頻度の単語を優先して回答してしまう機構に起因する。

- 誤答(質問応答の回答に対する質問)

質問応答の回答に含まれる用語について質問し、回答できなかった場合これに分類する。

- 誤答(感嘆としての発話)

驚きの「何それ! ?」など、質問以外の「何それ」について回答してしまった場合これに分類する。

- 誤答(質問タイミングが遅い)

質問のタイミングが遅かったために別の用語について回答してしまった場合これに分類する。

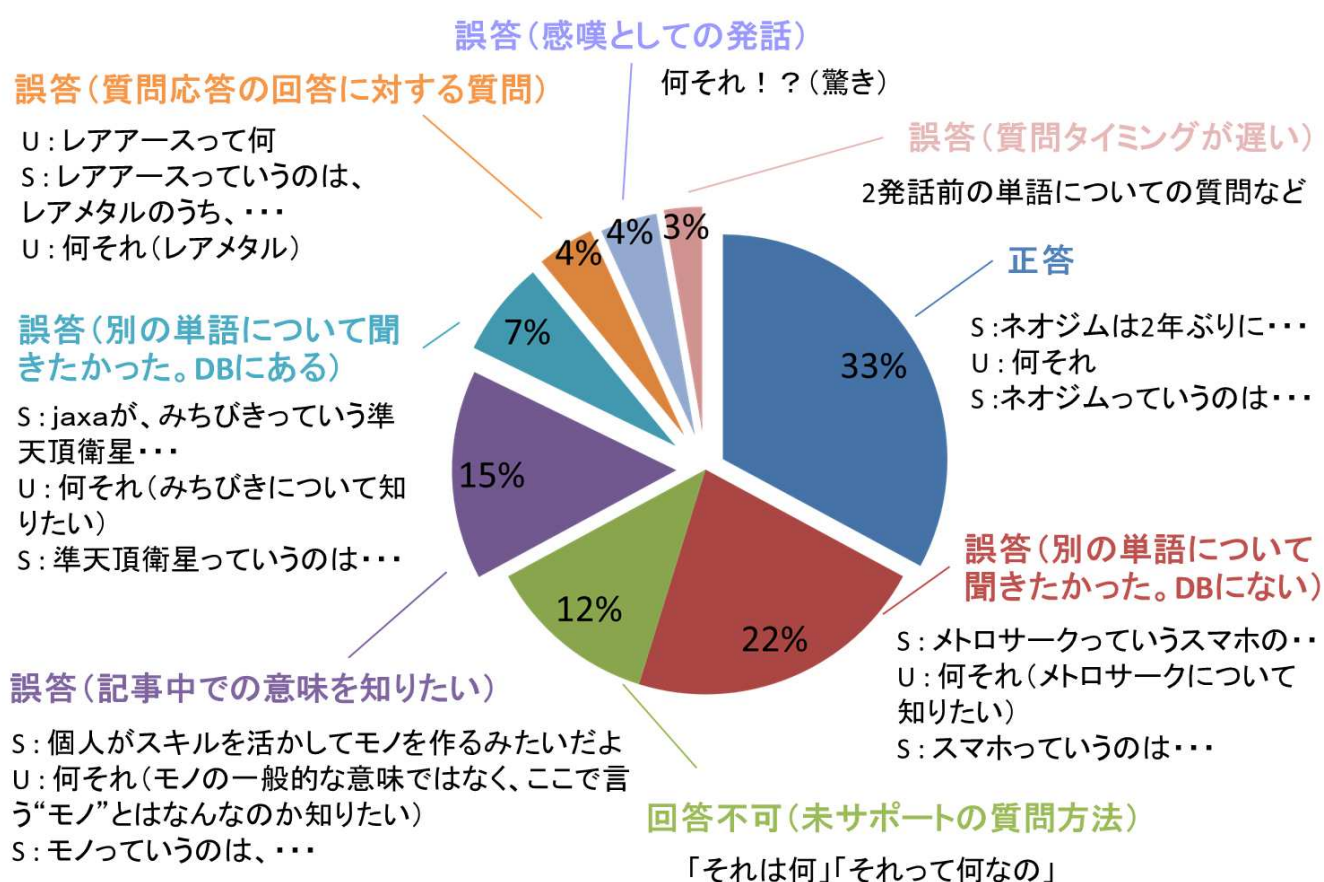


図 4.3: 「何それ?」に対するシステムの応答分類

次に、「○○って何?」の質問に対してシステムがどのような応答を行ったか以下の4つの項目に分類し、各割合を算出した結果を図 4.4 に示す。

- 正答

正しく回答できた場合これに分類する

- 回答不可 (DBにない)

質問対象の用語がデータベースになく「分からない」と回答した場合これに分類する。

- 誤答 (記事中での意味を知りたい)

記事中での意味を知りたかったが、一般的な定義を述べてしまった場合これに分類する。

- 誤答 (定義でなく、記事中の説明文を述べてしまった)

定義について知りたかったが、記事中の説明文を述べてしまった場合これに分類する。

記事中の説明を優先する機構に起因する。

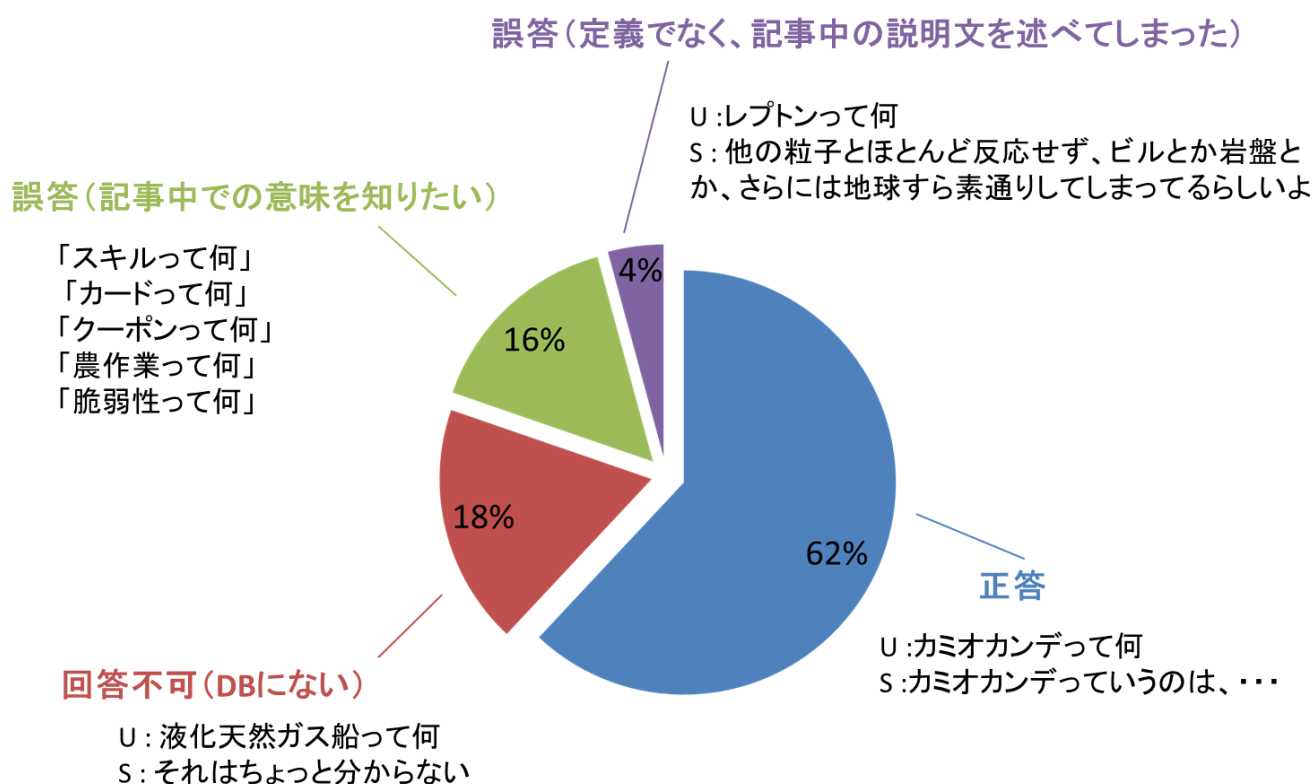


図 4.4: 「○○って何？」に対するシステムの応答分類

次に、その他の質問に対してシステムがどのような応答を行ったか以下の3つの項目に分類し、各割合を算出した結果を図 4.5 に示す。

- 正答

正しく回答できた場合これに分類する

- 回答不可

「分からない」と回答した場合これに分類する。

- 誤答

間違った回答をしてしまった場合これに分類する。

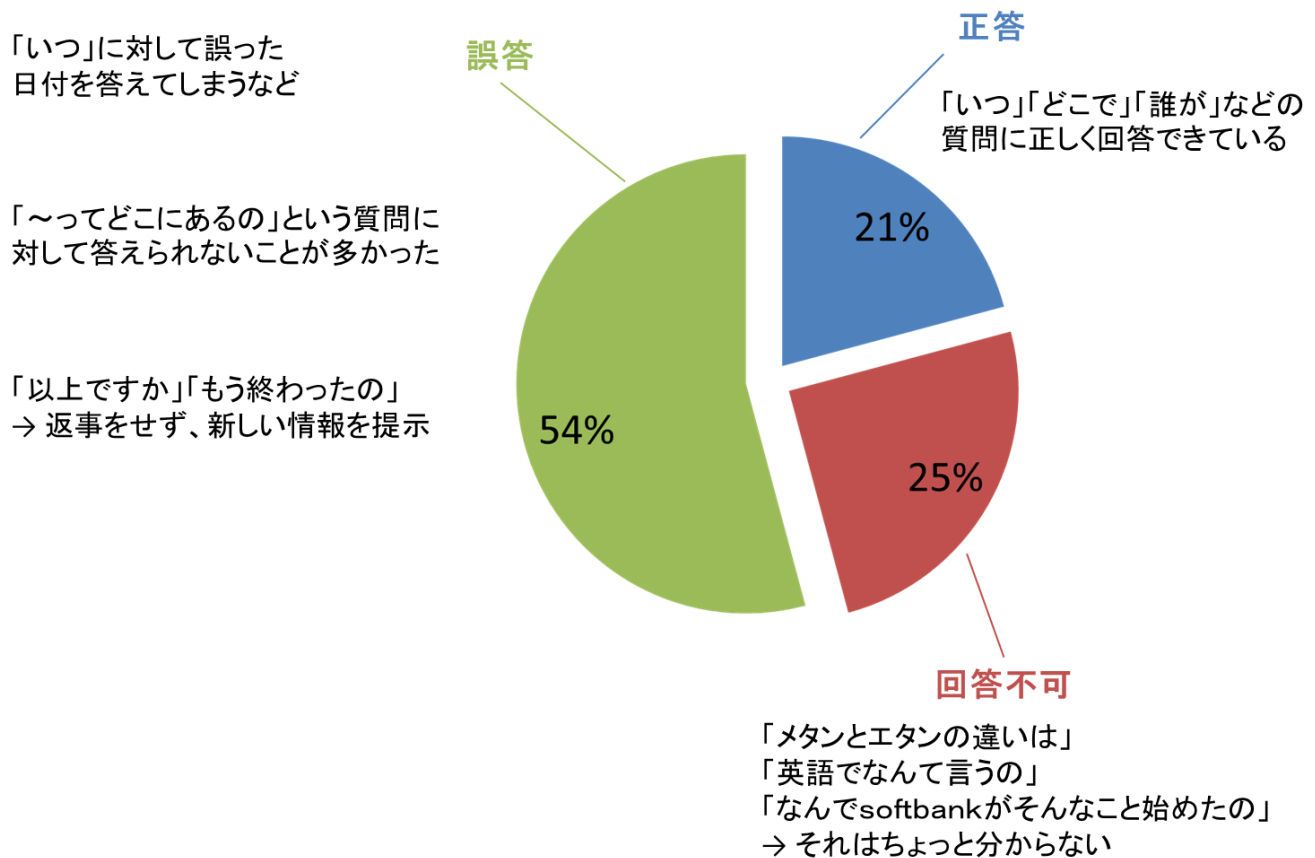


図 4.5: その他の質問に対するシステムの応答分類

4.3.3 応答能力の実力値

応答が求められるユーザー反応に対してシステムがどの程度適切に応答できたかを見積もる尺度として応答能力の実力値を定義する。応答能力の実力値は、各ユーザー発話タイプ（図 4.1）ごとに、会話で出現した割合とその正答率の積を計算し、和をとったものとする。ここで、「相槌・感想・その他」は応答義務が無いものとして正答率を 1 とする。また、音声認識結果が正しい「音声コマンド」も適切に動作すると考え、正答率を 1 とする。

「何それ？」は全体の 20% あり（図 4.1）、そのうち 33% が正しく応答できた（図 4.3）。「〇〇って何？」は全体の 19% あり（図 4.1）、そのうち 62% が正しく応答できた（図 4.4）。「その他の質問」は全体の 6% あり（図 4.1）、そのうち 21% が正しく応答できた（図 4.5）。音声認識結果が正しかった「音声コマンド」は全体の 15% あり、「相槌・感想・その他」は全体の 17% あった（図 4.1）。音声認識誤りがあった発話は全体の 23% あり、そのうち「相槌・感想・その他」が 41%、エラーハンドリングにより正答したものが 29% あった（図 4.2）。これらの結果から、シ

システムの応答能力を計算すると、実力値として68%の応答能力があることが分かった（図4.6）。

$$\frac{0.20 \times 0.33}{\text{何それ？}} + \frac{0.19 \times 0.62}{\text{〇〇って何？}} + \frac{0.06 \times 0.21}{\text{その他質問}} + \frac{0.15}{\text{音声}} + \frac{0.17}{\text{相槌}} + \frac{0.23 \times 0.7}{\text{音声認識誤り}} \div \mathbf{0.68}$$

正答 正答 正答 コマンド 感想 エラーハンドリング
その他

図 4.6: 現システムの応答能力の実力値

また、「何それ？」の質問方法のサポートを増やし、データベースを完備したと仮定すると、図4.3の「回答不可（未サポートの質問方法）」と「誤答（別の単語について聞いたかった。DBにない）」および、図4.4の「回答不可（DBにない）」の質問について回答できるようになるため、これにより応答能力は78%まで向上することが期待できる（図4.7）。

$$\frac{0.20 \times \mathbf{0.67}}{\text{何それ？}} + \frac{0.19 \times \mathbf{0.80}}{\text{〇〇って何？}} + \frac{0.06 \times 0.21}{\text{その他質問}} + \frac{0.15}{\text{音声}} + \frac{0.17}{\text{相槌}} + \frac{0.23 \times 0.7}{\text{音声認識誤り}} \div \mathbf{0.78}$$

正答 正答 正答 コマンド 感想 エラーハンドリング
その他

図 4.7: データベースを完備した場合の応答能力の理論値

今後は、データベースを充実させ回答できる質問の数を増やす一方で、一般的な用語について記事中での意味を回答する方法や、「何それ」以外の質問に対する回答率・回答精度を向上させる方法について検討する。

4.4 まとめと今後の課題

本章では、副計画の生成方法について説明し、システムの応答能力を評価した。

応答能力の実力値を定義し、「何それ」に特化したシステムとして評価したところ、現状のシステムの応答能力は68%程度であることが分かった。

今回は、質問応答の種類として、固有表現に関する質問応答、定義型質問応答、why型質問応答、真偽判定型質問応答を扱った。今後の課題として、これらの精度向上はもちろん必要であるが、how型質問や意見・評価を尋ねる質問など、より多様な質問に回答するための仕組みについても検討する必要がある。

用語の定義に関してはWikipediaなどの外部の情報を活用しているが、現在のシステムは、基本的に副計画は記事内の情報のみから作成している。今後、質問の回答率を向上させるには、関連記事などの情報も活用して副計画を拡充する必要がある。

また、副計画として質問応答だけでなく、ユーザーからのフィードバックを誘発するような、自己開示的な発話や主観的な意見などの対話戦略の組み込みについても検討する。

以下、今回扱った4タイプの質問応答についての課題を述べる。

固有表現に関する質問応答

固有表現抽出に KNP を使用しており、固有表現抽出精度がそのまま質問応答の回答率および正解率に直結している。また、現状、IREX の8種類の固有表現にしか対応していない。今後は、200種類からなる拡張固有表現 [Sekine 02, 関根 07, Sekine 08] に対応させるか、固有表現に頼らない回答抽出手法について検討する必要がある。

定義型質問応答

多義語に関する質問では、語義曖昧性解消を行ったうえで、回答する必要がある。例えば、現状のシステムでは、「NLP って何?」という質問があったとき、よく用いられる意味として、「自然言語処理 (Natural Language Processing)」の略称として説明するが、文脈によっては、「非線形計画法 (Non-Linear Programming)」の略称であったり、「神経言語プログラミング (Neuro-Linguistic Programming)」の略称であったりする。そのため、文脈に応じて語義を特定する必要がある。

why 型質問応答

ニュース記事は、過去に配信されたニュースの内容を既知として、その出来事の背景や経緯などについての説明が省略されていることがある。現在は、記事内の情報のみしか活用していないが、今後は、関連する過去の記事も活用し、理由や背景を抽出する方法について検討する。

真偽判定型質問応答

ここでは、データベースの知識を用いてマッチングに基づく方法で述語項構造間の含意矛盾認識を行った。近年、含意関係認識 (Recognizing Textual Entailment) の分野では、深層学習の手法が主流になりつつある [Devlin 18, Rocktaschel 16, Wang 16]。今後は、このような手法の有効性についても検証したい。

第5章 ユーザー発話の意図理解

5.1 序論

我々はニュース記事のようなまとまった量の情報を効率的に伝達する会話システムの開発を行っている。ここで「効率的」とは、伝達対象となる記事の中から、ユーザーにとって不要な情報を除き、必要な情報だけを伝えることを意味する。我々のシステムの特徴は、あらかじめ主計画、副計画と呼ぶ複数のシナリオを用意しておき、このシナリオに沿って会話を進めることで、リズムの良い会話を実現するうえで必須となる迅速な応答を可能としたところにある。主計画に沿って記事の要点となる情報を提示する傍らで随時ユーザーからのフィードバックを理解し、必要に応じて副計画に遷移して補足情報を提示する。このようにユーザーの興味や理解状態に応じて提示する情報を柔軟に切り替えながら会話を進めていく仕組みを持つ。一方で、高い情報伝達効率を実現するには、ユーザーからのフィードバックを正しく理解することが重要となる。

フィードバックの役割は、ユーザーからシステムにある種の要求を伝えることにある。要求は、ユーザーが直接明示的に伝えるものと、ユーザー状態などを通して間接的あるいは暗示的に伝えられるものがある。本研究では、これらをまとめて発話意図と呼ぶ。

我々は、システムが伝える情報量の観点から、その増減を要求する発話意図を分類した [Yokoyama 18]。伝達情報の増加を求める発話意図として「質問」「補足要求」「反復要求」を、伝達情報の減少を求める発話意図として「無関心」「既知」を定めた。加えて、快適に情報の授受を行うためには、ユーザーの発話とシステムの発話が同時に起こるべきでないと考え、発話衝突の回避を求める発話意図として「待機要求」を定めた。

本研究では、これらの発話意図を識別するモデルを次の観点で設計する。

(1) スペクトログラムからの韻律情報の抽出

ユーザーのフィードバックは必ずしも言語的に明示された形で表れるとは限らない。場合によっては、抑揚で表現されるニュアンスなどにユーザーの意図が表れることもある。従来、感情認識などのタスクでは特徴量として基本周波数 (F0) が用いられてきた。しかしながら、音声波形の準周期性や周辺雑音、有声音中の基本周波数の変化が広域に渡るなどの理由により、

基本周波数を正確に抽出するのは難しい。そこで、F0の推定を介さずに音声の時間・周波数スペクトルから直接特徴量を抽出する方法を提案する。具体的には、音響情報としてスペクトログラムをCNNを含むAutoEncoderに入力し、その中間層出力を韻律特徴量として用いる。

(2) システム発話文脈の考慮

フィードバックに込められる意図（以下、発話意図）をユーザーの発話情報のみから判断するのは困難である。例えば、「え？」という発話が驚きなのか、質問なのか、聞き返しなのかは文脈に依存する。そこで、本研究では、ユーザー発話から抽出した言語特徴量だけでなく、システム発話を文脈として考慮した発話意図認識モデルを提案する。さらに、言語情報のエンコーダーにBERT (Bidirectional Encoder Representations from Transformers) [Devlin 18]を採用したうえで、Masked Language ModelとNext Sentence PredictionによるBERTの事前学習の発話意図認識における効果についても確認する。

本章の構成は次の通りである。5.2節で関連研究について述べる。5.3節で発話意図データセットについて説明し、5.4節で提案する発話意図認識モデルについて説明する。5.5節で発話意図データセットを用いてモデルを評価した結果を報告する。最後に、5.7節でまとめと今後の課題について述べる。

5.2 関連研究

従来、発話意図認識や感情認識のタスクでは、人手で設計した特徴量が利用されてきた [Ando 15, 藤江 03, 岩田 12, 林 14, Nisimura 06, 田中 98]。例えば、藤江らは、システムに対する利用者の発話態度（肯定的か否定的か）を推定するために、第1モーラの基本周波数 (F0) の傾き、発話全体のF0レンジ、最終モーラの継続長からなる3次元の特徴量を使用した [藤江 03]。Andoらは、対話データにおいてその発話が肯定的であるか否定的であるかを推定するために、音響特徴量として単語ごとに算出したF0、パワーの最大・最小、継続長、間などの情報を、言語特徴量としてバイグラム言語モデルのパープレキシティを使用した [Ando 15]。Nisimuraらは、収集した子供の音声に対して喜んでいるか嫌悪を示しているかの推定を行う際、F0値、パワーから求められる16次元の特徴量の中から、因子分析により寄与の高い因子を選択することで、より識別率の高い特徴量を使用している [Nisimura 06]。田中らは、人手によってテキストのみから判断した文型（疑問文、平叙文、命令文）に加え、発話末F0値の時間的变化から判別した音調パターンを特徴量としている [田中 98]。林らは、発話タグ（疑問文、平叙文、相槌、同意、

笑い)の推定において、F0 値やパワー、話速などの 23 次元の音響特徴量を使用した [林 14]。岩田らは、文末詞とその音調との組み合わせにより、聞き手に伝わる発話意図を分析している。その際、発話末の F0 値を分析し、近似、正規化などを行った F0 形状に対してクラスタリングを行っている [岩田 12]。これらの研究は、韻律情報を詳細に観察・分析することで意図理解のための特徴パラメータを設計している。

しかしながら、人手で設計可能な特徴パラメータは、推定対象の発話意図の違いが文章や音声を観察することで見いだせる場合に限られる。つまり、観察・分析だけでは容易に特徴が見つけられないような微妙なニュアンスの意図については特徴量の設計が困難になる。加えて、発話意図ごとに特徴パラメータの設計が必要となる。

近年では、スペクトログラムを直接入力して感情認識などを行う研究が増えてきている [Guo 18, Luo 18, Satt 17, Tang 18, Yenigalla 18]。例えば、Guo らは、スペクトログラム、位相情報、MGDCC [Hegde 07] を CNN に入力して得られた特徴量を双方向 LSTM に与え感情（怒りや喜びなど）を識別するモデルを提案した [Guo 18]。Luo らは、スペクトログラムを CRNN に入力して得られた特徴量と人手で設計された特徴量（F0 や MFCC など）を組み合わせる感情を識別するモデルを提案した [Luo 18]。Yenigalla らは、スペクトログラムを CNN に入力して得られた特徴量と word2vec [Mikolov 13a, Mikolov 13b, Mikolov 13c] で得られた音素の埋め込み表現を組み合わせる感情を識別するモデルを提案した [Yenigalla 18]。

本研究では、スペクトログラムを CNN を含む AutoEncoder に入力し、その中間層出力を音響特徴量として LSTM に与える。そして、その LSTM の情報とユーザー発話の言語情報に加え、直前のシステム発話の言語情報を考慮した発話意図認識モデルを提案する。

5.3 発話意図データセット

発話意図の情報が付与されたコーパスとして、横山らが構築したデータセットを使用した [Yokoyama 18]。このデータセットは、我々が情報伝達のために開発した即応性に富む会話システムと 24 名の大学生が会話して得られた約 2,000 対話分の音声対話データに基づいて作られたデータセットである。収集したユーザー発話のうち、VAD で切り出した 1.5 秒以下の音声に対して 10 人のアノテーターが発話意図に関するアノテーションを行った。

発話意図の分類と対応するシステム動作を表 5.1 に示す。伝達情報の増加を求める発話意図として「質問」「補足要求」「反復要求」を、伝達情報の減少を求める発話意図として「無関心」「既知」を、発話衝突の回避を求める発話意図として「待機要求」を定めた。本研究の実験で利用した発話意図データセットの統計を表 5.2 に示す。

アノテーター間のラベルの一致率を図 5.1 に示す。「質問」以外のラベルの一致率が低い理由として、アノテーターの教育不足およびフィードバックの質が悪いことが考えられる。後者に関しては、質問応答や音声合成の質が十分でないこと、ユーザーとシステムが良好な関係を築けていないことなどが原因として考えられる。

表 5.1: 発話意図の分類と対応するシステム動作

効果	発話意図	システムアクション	発話例
伝達情報を増やす	質問	質問応答	誰が？，どこで？，何それ，ネオジム？
	補足要求	補足要求	ん？，へー！，そうなんだ！，すごいね
	反復要求	繰り返し	もう一回言って，え？ 何？，ん？なんだって？
伝達情報を減らす	無関心	他のトピックへの以降	ふーん，へー，あー，そっか，そうなんだ
	既知	詳細説明の省略	それ知ってる，あーその話，うんうん
発話の衝突を避ける	待機要求	傾聴	それってさ，いや，えっと，じゃ糸魚川の…

表 5.2: 実験で使した発話意図データセットの統計

	訓練セット		開発セット		テストセット	
	正例	負例	正例	負例	正例	負例
質問	2042	2042	292	276	584	584
補足要求	3228	3197	461	457	924	915
反復要求	360	358	52	52	102	102
無関心	1776	1763	254	252	508	504
既知	355	350	51	51	101	102
待機要求	361	359	52	52	104	102

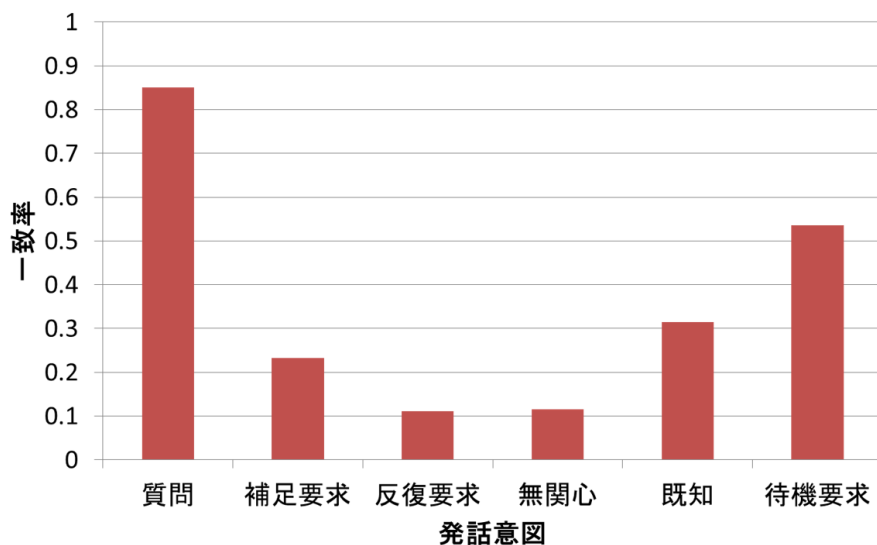


図 5.1: 発話意図ごとのラベラー間のラベル一致率 (Fleiss' Kappa)

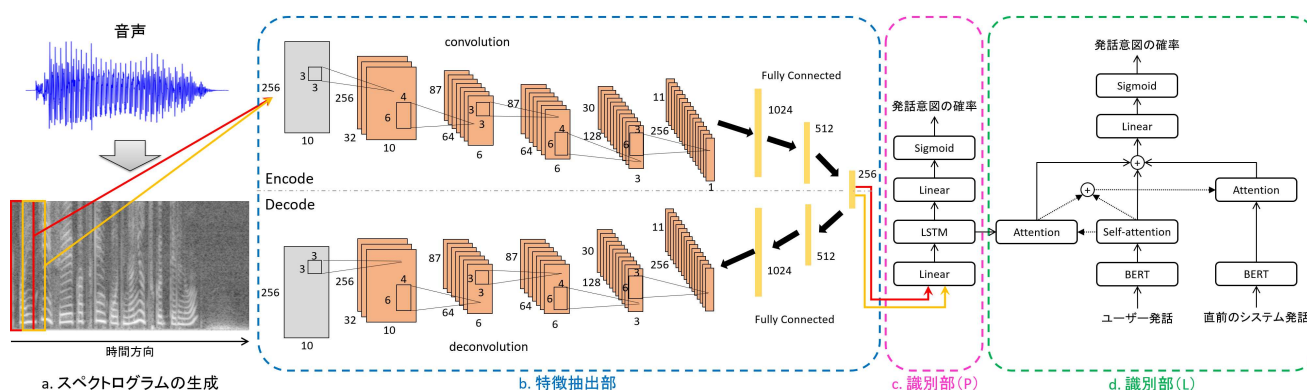


図 5.2: 発話意図認識モデル

5.4 発話意図認識モデル

提案する発話意図認識モデルの全体像を図 5.2 に示す．まず，短い時間幅で切り出した音声の断片からスペクトログラムを生成する．次に，得られたスペクトログラムを CNN を含む AutoEncoder (CNN-AutoEncoder) に入力し，その中間層に圧縮された韻律特徴量を時系列に沿って LSTM に入力する．LSTM は逐次発話意図ラベルの確率を出力するが，音声認識結果が得られた段階で，LSTM の隠れ層の情報と，ユーザー発話の言語情報および，直前のシステム発話の言語情報を統合して最終的な発話意図の推定結果を得る．以下，特徴抽出部と識別部について説明する．

5.4.1 特徴抽出部の設計

一般に発話意図の推定では、特徴量として基本周波数 (F_0) が用いられる。しかしながら、音声波形の準周期性や周辺雑音、有声音中の基本周波数の変化が広域に渡るなどの理由により、基本周波数を正確に抽出するのは難しい。そこで、 F_0 の推定を介さずに音声の時間・周波数スペクトルから直接特徴量を抽出する方法を提案する。

音韻や声の高さに関する特徴はスペクトログラムの模様として表れる．そこで，このスペクトログラムの模様を二次元の画像と見なして，5層の畳み込み層と3層の全結合層からなるネットワークを折り返した全16層で構成されるCNN-AutoEncoder（図5.2.b）を学習する．そして，その中間層に圧縮された韻律特徴量を発話意図の識別で利用する．

5.4.2 識別部の設計

識別部は、韻律情報のみから発話意図を識別する識別部（P）と、識別部（P）で得られる韻律情報に加え、ユーザー発話および直前のシステム発話の言語情報を考慮して発話意図を識別する識別部（L）から構成される。

5.4.2.1 識別部（P）

識別部（P）では、CNN-AutoEncoder から取り出された韻律特徴量を LSTM に逐次入力して発話意図を識別する（図 5.2.c）。

音声は時間方向に可変長であり、時間方向の長さに頑健なモデルであることが望まれる。同じ文字列の音声でも人の違いや発話する条件や状態によってその継続長は異なる。また、発話末のピッチが上昇すると「質問」と捉えやすくなるなど、発話意図認識において韻律の時間方向の変化は有用な情報である。本研究では、RNN の中でも長期の依存性に長けた LSTM を用いた。

5.4.2.2 識別部（L）

識別部（P）の LSTM の状態とユーザー発話および直前のシステム発話の言語情報を用いて発話意図を識別する（図 5.2.d）。ここで、直前のシステム発話は、ユーザー発話を受け付けた時点での現在の発話内容とその一つ前の発話内容を表す。例えば、表 5.3 のような会話において、ユーザー発話「何それ」の発話意図を識別する場合、文脈として用いるシステム発話は現在の発話内容「ツイートにスクリプトを」とその一つ前の発話内容「攻撃は XSS の脆弱性を突いたものだって」となる。

表 5.3: 文脈として用いるシステム発話の範囲（太字部分）（S がシステム発話，U がユーザー発話を表す。）

S :	ツイッターのアカウントがハイジャックされる 事件が起きたらしいよ
S :	攻撃は XSS の脆弱性を突いたものだって
S :	ツイートに スクリプトを
U :	何それ

ユーザー発話および直前のシステム発話の言語情報のエンコードには, BERT [Devlin 18] を用いた. BERT は, Transformer [Vaswani 17] の Encoder 部分をユニットとする双方向 Transformer モデルである. 文の単語をランダムにマスクし, そのマスクされた単語を予測する Masked Language Model と二つ文が隣接しているかどうかを予測する Next Sentence Prediction の2つのタスクで事前学習したモデルを転移学習させることで, 自然言語処理の様々なタスクで SOTA を達成し, 汎用的な言語表現を獲得できるモデルとして注目されている. 本研究では, BERT を発話意図認識タスクに適用し, 事前学習の効果を検証する.

ユーザー発話の音声認識結果を BERT でエンコードした結果に対して self-attention [Lin 17] を適用し, 得られたベクトル \mathbf{v}_u をクエリとして韻律情報を保持する LSTM の各状態に対して attention [Bahdanau 15] を計算する. その結果得られる文脈ベクトル \mathbf{v}_p とユーザー発話のベクトル表現 \mathbf{v}_u を結合したベクトルをクエリとして, システム発話のエンコーダー BERT に対して attention [Bahdanau 15] を計算する. その結果得られる文脈ベクトル \mathbf{v}_s を \mathbf{v}_u および \mathbf{v}_p と結合したベクトルを出力層に与え, 最終的な発話意図の確率を求める.

5.5 発話意図識別実験

5.5.1 特徴抽出部の学習と結果

特徴抽出部 (図 5.2.b) の CNN-AutoEncoder の学習には, 大規模な『日本語話し言葉コーパス (CSJ)』¹を用いた.

CNN-AutoEncoder の入力, フレームサイズ 800 (50ms), フレームシフト 160 (10ms), チャンクサイズ 1024 で切り出した音声から生成したスペクトログラムを時系列に並べたものとし, そのサイズは 10×256 とした. このデータをもとにネットワークの学習を行い, 特徴抽出器を構築した.

学習した CNN-AutoEncoder の出力例を図 5.3 に示す. ここで, 中央のスペクトログラムが入力として与えたスペクトログラムで, 右側のスペクトログラムが CNN-AutoEncoder が出力したスペクトログラムである. 左側のグラフは各スペクトログラムの 5 番目のスペクトルの周波数強度を表している. これらの結果からモデルが入力スペクトログラムの模様パターンを復元できていることが分かった.

¹http://pj.ninjal.ac.jp/corpus_center/csj/

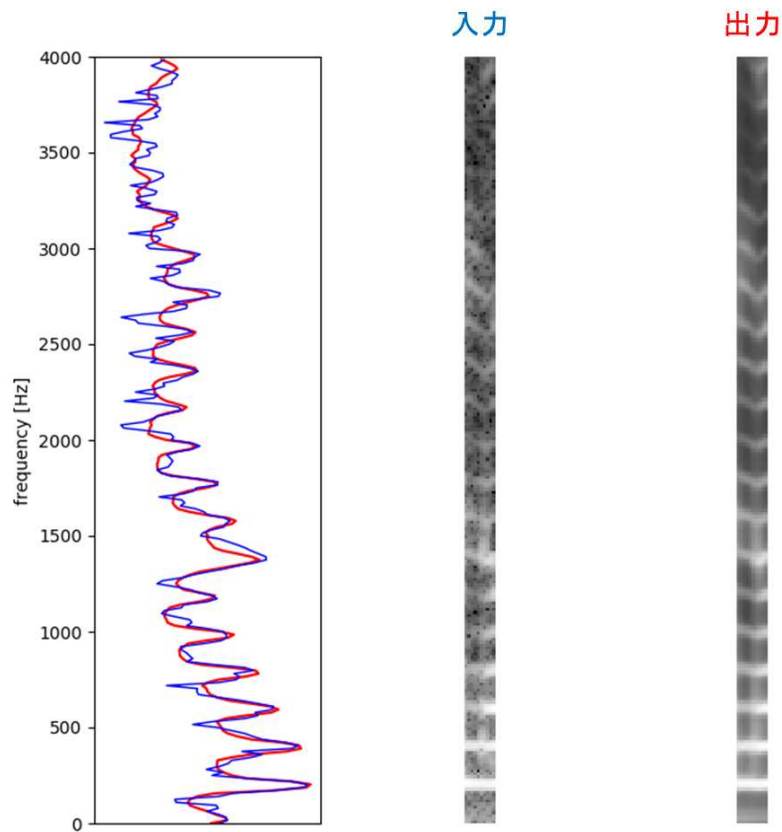


図 5.3: CNN-AutoEncoder の出力例 (左:各スペクトログラムの5番目のスペクトルの周波数強度 (青:入力, 赤:出力), 中:入力したスペクトログラム, 右:出力したスペクトログラム)

5.5.2 識別部の実験設定

Linear 層および LSTM の隠れ層の次元は 64 に設定した．BERT の事前学習の詳細および補助情報として用いた言語特徴量の説明を以下に示す．

5.5.2.1 BERT の事前学習

日経新聞の 200476 個のニュース記事から段落を超えないように隣接文ペアを重複なく抽出した．このうち, 700000 文ペアを訓練セット, 37094 文ペアを開発セットとして, Masked Language Model と Next Sentence Prediction の 2 タスクで BERT の事前学習を行った．語彙には訓練セットにおいて頻度が 7 以上であった 63272 語を用いた．モデルのパラメータは, Transformer のブロック数を $L = 8$, 隠れ層の次元を $H = 256$, self-attention のヘッド数を $A = 8$ に設定した．

5.5.2.2 補助情報

システム発話の補助情報には、JUMAN++² (Ver.1.02) の形態素情報 (品詞大分類, 品詞細分類, 活用形, 活用型, カテゴリ, ドメイン), 単語の TF, IDF, TF-IDF, 「」『』内の単語かどうか, KNP³ (Ver.4.19) を適用して得られる IREX の8種類の固有表現クラス, 係り受けの種類, 係り受け木の深さ, 係り元の文節数, 文頭からの文節位置を用いた。

シナリオにあるシステム発話は事前に解析しておくことができるのに対し, ユーザー発話はリアルタイムに解析する必要がある。そのため, ユーザー発話の補助情報には, JUMAN⁴ (Ver.7.01) を適用して得られる形態素情報のみを使用した。

5.5.3 実験設定

発話意図データセット (5.3 節) を用いて, 識別部 (P) を学習し, LSTM の最終状態の出力結果をもとに, 韻律情報のみを用いたときのテストセットに対する Accuracy を計算した。また, 学習済みの識別部 (P) の LSTM の隠れ層の値とユーザー発話 (書き起こし) およびシステム発話の言語情報を用いて, 識別部 (L) を学習し, テストセットに対する Accuracy を計算した。

韻律情報のみを用いたとき (P) と, これにユーザー発話の言語情報を加えたとき (P+U), さらにシステム発話の言語情報を加えたとき (P+U+S) の比較を行った。また, BERT を事前学習したときとしなかったとき, 補助情報を加えたときと加えなかったときについても比較を行った。

5.5.4 実験結果

実験結果を表 5.4 に示す。全体の傾向として, 韻律情報のみよりもユーザー発話の言語情報を加えたときの方が良く, さらに直前のシステム発話の言語情報を加えることで Accuracy が向上することが分かった。また, BERT を事前学習することで, 補助情報を加えることで Accuracy が向上することが分かった。

しかしながら, 「無関心」の識別に関してはシステム発話の情報を加えることで性能が悪化した。これは, コンテンツに対する興味の傾向がユーザーごとに様々であることに起因すると考えられる。そのため, 今後は, ユーザーを区別する識別子を補助情報に加えるなどして改善を図る。

²<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

³<http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

⁴<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

表 5.4: 発話意図認識の実験結果

	事前学習あり+補助情報あり			事前学習あり+補助情報なし		事前学習なし+補助情報なし	
	P	P+U	P+U+S	P+U	P+U+S	P+U	P+U+S
質問	0.902	0.955	0.960	0.932	0.944	0.929	0.935
補足要求	0.687	0.753	0.759	0.699	0.703	0.697	0.697
反復要求	0.593	0.775	0.789	0.618	0.637	0.598	0.613
無関心	0.760	0.779	0.772	0.770	0.768	0.768	0.760
既知	0.552	0.749	0.759	0.709	0.734	0.571	0.606
待機要求	0.709	0.738	0.752	0.728	0.733	0.714	0.728

5.6 発話意図「質問」に関する予備実験

ラベラー間の一致率（図 5.1）が高くデータ数が多い発話意図「質問」に関して識別実験を行った。この実験では、システム発話の言語情報は使用せず、韻律情報として LSTM の最終状態の隠れ層の値を用いた。また、ユーザー発話の言語情報は Bag-of-Words で与えた。これらのベクトルを結合したものを SVM などの識別器に与え、発話意図の識別を行った。

5.6.1 「質問」識別用データセット

本実験で用いた「質問」データセットの統計を表 5.5 に示す。ここで、「質問」以外の発話意図に分類された発話を「非質問」とする。

表 5.5: 「質問」「非質問」データセットの統計

	総数	「質問」の数	「非質問」の数
訓練セット	2,000	925	1,075
テストセット	1,000	467	533
開発セット	257	138	119

5.6.2 LSTM の学習と識別結果

まず、5.5.1 節で学習した CNN-AutoEncoder の中間層の値（256 次元）を入力として LSTM の学習を行った。ここで、CNN-AutoEncoder への入力には、100ms のスペクトログラムを 50ms ずつシフトさせながら与えた。つまり、現時刻の入力と次の時刻の入力で 50ms のオーバーラッ

プが存在する。これは、LSTM に時間変化の情報を陽に学習させるためである。モデルのパラメータは、入力層のユニット数を 100、中間層のユニット数を 50、出力層のユニット数を 50 に設定した。

LSTM の最終層の出力ラベルと正解ラベルの一致率を求めたところ、Accuracy は 0.896、「質問」を「質問」と識別する精度は 0.878、「非質問」を「非質問」と識別する精度は 0.912 であった。エラー分析を行ったところ、「なにそれ」や「どうして」「誰が」など、言語表記を見れば「質問」と判別できるような発話を「非質問」と判定してしまう事例が多々見られた。このような誤りを解消すべく、次節では発話の言語情報も組み合わせた発話意図の識別について検討した。

5.6.3 韻律情報と言語情報を用いた識別器の学習と識別結果

本実験では以下の観点について検証を行った。

- 有効な識別器

scikit-learn⁵の識別器を一通り比較し、有効な識別器について調査する。各識別器のパラメータには初期値を用いた。

- 言語情報として有効な素性の組み合わせ

言語情報には、基本素性としてユーザー発話の文字または単語の Bag-of-Words (BoW) を使用し、追加素性として JUMAN (Ver.7.01)⁶を適用して得られる形態素情報（品詞大分類、品詞細分類、活用形、活用型、カテゴリ、ドメイン）を使用した。これらの素性に関して有効な組み合わせについて調査する。

- 言語情報のみ、韻律情報のみ、韻律情報と言語情報の3通りで識別器を学習したときの比較

識別器を言語情報のみを用いて学習した場合と韻律情報のみを用いて学習した場合、韻律情報と言語情報を組み合わせて学習した場合で性能を比較し、各特徴量の有効性を確認する。韻律情報には 5.6.2 節で学習した LSTM の最後の隠れ層の値を用いた。

- 発話の表記として人手で書き起こした結果を用いたときと音声認識結果を用いたときの比較

人手で書き起こした結果から抽出した言語特徴量を用いた場合と音声認識結果から抽出した言語特徴量を用いた場合でどの程度性能に差が見られるか確認する。音声認識には Google Cloud Speech API⁷を使用した。

⁵<http://scikit-learn.org/stable/>

⁶<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

⁷<https://cloud.google.com/speech/?hl=ja>

- ユーザー発話の言語表記に関してルールで識別したときと機械学習で識別したときの比較

ベースラインとして基本的なルールでどの程度の識別性能を実現できるか確認する。実際に用いたルールを以下に示す。

- － 以下を含めば「質問」
「何」「なに」「なにに」「どこ」「だれ」「誰」「なんで」「なんの」「なん日」「どう」「いつ」「いくら」「なぜ」「どんな」「どの」「どれ」「例えば」「たとえば」「なんて」
- － 以下で終われば「質問」
「か」「かね」「かな」「の」「って」「つけ」「は」「で」「が」「に」「ってこと」
- － 以下に一致していれば「質問」
「え」「ん」「えっ」
- － 以下を含めば「非質問」
「そっか」「そうか」「なんでもない」「何でもない」

これらのルールは訓練セットの書き起こしデータに基づいて設計した。

実験結果を図 5.4 に示す。まず、言語情報のみを用いてルールで識別した場合と機械学習で識別した場合について比較すると、ルールによる識別では「質問」を「質問」と識別する性能が低いことが分かった。次に韻律情報のみを用いたときと言語情報のみを用いたときの結果を比較すると、言語情報のみ（機械学習）の方が高い性能を示した。

最も良い結果を示したのは、韻律情報と言語情報（単語の表層形と活用形）を特徴量に用いて線形カーネルの SVM を学習し識別を行ったときであった。ここで、書き起こしではなく音声認識結果を用いたときの Accuracy の劣化は 0.7%にとどまっていた。誤り例を表 5.6 に示す。書き起こしを用いたときは正しかったが、音声認識結果を用いたときに誤った事例について確認したところ、音声認識できなかったもの（e.g. 「えっ」→「」）や音声認識誤りで疑問詞が消えてしまったもの（e.g. 「え何年」→「宇奈根」, 「え何」→「えなり」）が誤りとして多く見られた。また、言語的に誤りを含むものだけでなく、言い淀みなどで音声的に不完全な発話も誤りやすいことが分かった。

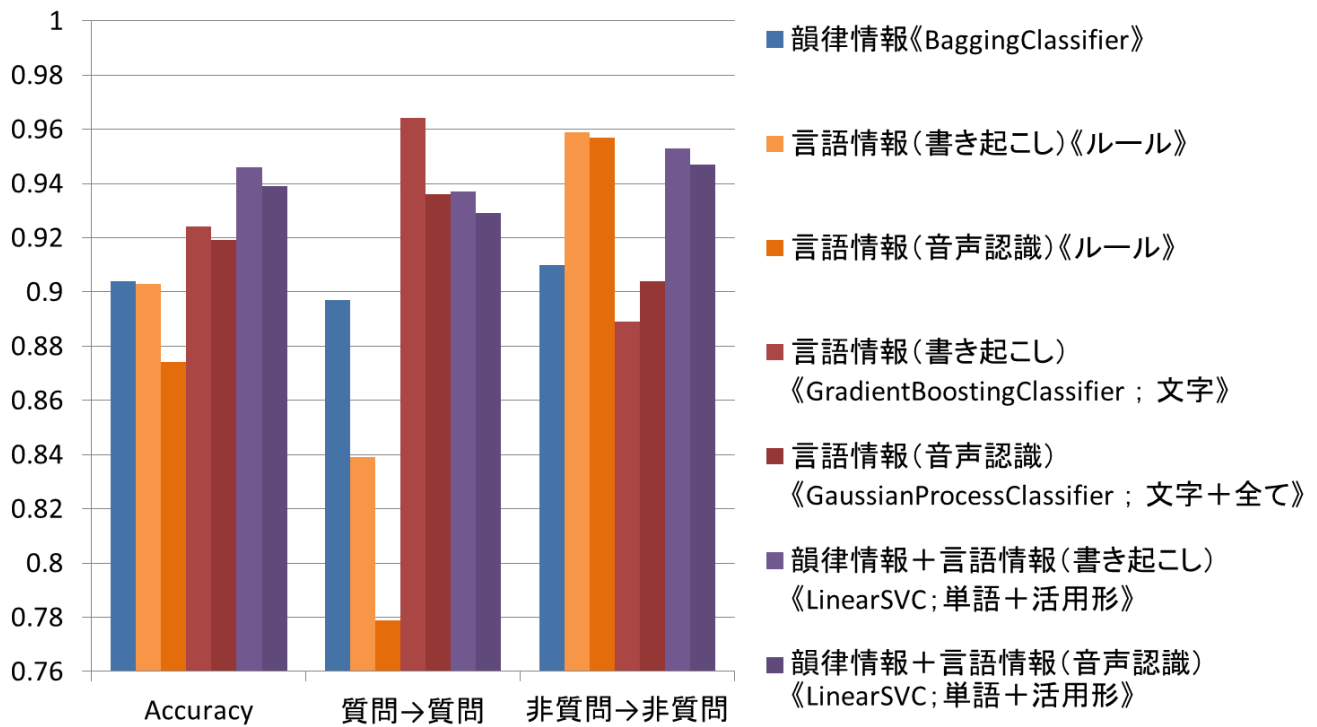


図 5.4: 発話意図「質問」の識別性能

表 5.6: 識別誤りの例

正解	出力	書き起こし	音声認識結果
質問	非質問	えっ	
		てなに	
		え何年	奈良根
		何ドルに	マンドリン
非質問	質問	低調	
		おー	
		んなんじゅ	なん j
		の	の

5.7 まとめと今後の課題

システム発話の文脈を考慮した発話意図認識モデルを提案し、直前のシステム発話の言語情報を文脈に用いることで発話意図認識の精度が向上することを確認した。

また、ユーザー発話およびシステム発話の言語情報のエンコーダーに BERT を採用し、Masked Language Model と Next Sentence Prediction の 2 タスクで事前学習することで、精度が向上することを確認した。

今後は、ユーザーの違いを考慮できるように特徴量を工夫するなどして性能の改善を目指すとともに、表情などの視覚情報も考慮したマルチモーダルな発話意図認識手法についても検討する。

また、発話意図のアノテーションに関して、アノテーター間のラベルの一致率が低い問題については、次のような観点も含め、より詳しい調査が必要である。

- (1) システムの使用感が改善されれば、良質なフィードバックが得られ、一致率が向上するか。
- (2) アノテーター間でラベリング基準がしっかりと共有されれば、一致率が向上するか。
- (3) 音声だけでなく、表情も考慮してアノテーションを行うことで、一致率が向上するか。

第6章 音声合成と間(ま)の制御

6.1 序論

ニュースのようなまとまった量の情報を伝える会話システムにおいて、談話構造上の発話の役割に応じたメリハリのある話し方を可能とする音声合成手法を提案する。

Siri¹や Pepper²といった音声対話システムの登場に伴い、人とシステムとのインタラクションのあり方が変化してきている。古くは、ユーザーの質問に応じてシステムが短い回答を提示する質問応答型の対話システム [Seneff 00, 駒谷 01, 駒谷 03a, 駒谷 03b] が主流であったが、近年は、主体的にまとまった情報を提供する対話システム [吉野 14, Yoshino 15a, Yoshino 15b, Traum 15a, Traum 15b, Traum 16] が提案されてきている。

我々も、ニュース記事に代表されるまとまった情報をユーザーとの会話インタラクションを通じて効率的に伝達する会話システムの開発を行っている。このシステムでは、あらかじめ主計画、副計画と呼ぶ複数のシナリオを用意したうえで、主計画に沿って記事の要点となる情報を提示する傍らで随時ユーザーからのフィードバックを理解し、必要に応じて副計画に遷移し補足情報を提示する。この機能によって、情報提示と質疑応答を頻繁に、かつ迅速に切り替えながら効率的に情報を伝えることができる。

このようなまとまった量の情報を伝える会話システムでは、重要な情報がユーザーに伝わる話し方が求められる。話し方の教則本などでもよく指摘されているように、情報を伝える話し方として、重要な箇所をゆっくり話して強調する、相手に理解やフィードバックを促すように適度に間(ま)を置くなど、メリハリをつけて話すことが重要である [西野 12, のぎき 18]。この発話のメリハリの問題は、音声合成の分野では、これまで一文内におけるプロミネンスの付与という形で扱われてきた [Malisz 17, 武田 91a, 武田 91b]。しかしながら、発話の強調に影響を与える韻律は、談話構造に応じて変化することが古くから指摘されており [Hirschberg 96, 岩田 94, Koiso 97]、一文より長い単位で韻律や間(ま)を制御する必要がある。

本研究では、このメリハリを実現するために発話の談話構造上の役割に着目する。音声合成の基盤システムとして DNN 音声合成システム Merlin [Wu 16] を採用したうえで、談話構造上の文の役割を、核(最も伝えたい内容を含む文)、前置き(核を伝えるためのリード文)、補足

¹<http://www.apple.com/jp/ios/siri/>

²<https://www.softbank.jp/robot/consumer/products/>

（情報を補足する文）の3つに分類し、これらを DNN 音声合成の補助情報として利用することで、発話系列全体で韻律や間（ま）の制御が可能な音声合成手法を実現する。

本章では、まず、6.2 節で関連研究について述べた後、6.3 節で提案システムの構成について説明する。次に、6.4 節で音声合成およびポーズ長（＝間（ま）の長さ）推定のための音声収録の方法について述べ、収録した音声の特徴を述べる。そして、6.5 節で音声合成手法およびポーズ長推定手法について説明した後、6.6 節、6.7 節、6.8 節、6.9 節で提案手法の有効性を評価した結果について報告する。最後に、6.12 節でまとめと今後の展望について述べる。

6.2 関連研究

ニュースの情報伝達に近い分野として、Storytelling を対象とした音声合成の研究がある。Theune らは、人の語り口調の音声を収集し、物語において緊張感のある場面を語る際の発話の韻律変化に着目し、それを制御する合成器を開発した [Theune 06]。Montano らは、語り口調、読み上げ、対話などの発話スタイルや、キャラクター口調、緊張感のある状況といった語り口調における調子の異なる音声を収録し韻律の分析とそれを反映した合成音声の作成を行った [Montano 13]。Sarkar と Verma らは、プロの Storyteller に発話させたヒンディー語の音声データを収集し、ピッチやテンポ、強さやポーズについて分析を行い、ニュートラルな発話スタイルを Storytelling の発話スタイルに変換するルールを定めた [Sarkar 14, Verma 15]。Ramli らも、マレー語の Storytelling データに関して同様の分析を行い、ニュートラルな発話スタイルを Storytelling の発話スタイルに変換するルールを定めた [Ramli 16, Ramli 17]。

Storytelling におけるポーズに焦点を当てた研究も行われている。Sarkar らは、Storytelling におけるポーズが物語の緊張感を保つことや、注目を集める役割を果たすとして、決定木を用いたポーズ位置の予測を行った [Sarkar 15a]。また、後続研究ではポーズのパターンを short, medium, long の3種類に分類したうえで、ポーズ長の予測を行った [Sarkar 15b, Sarkar 15c]。

このように物語を対象として、一文より長い単位での音声合成の研究は行われているものの、これらは感情の使い分けや発話スタイルの変換を対象としたものとなっており、重要な情報を印象付けるようなメリハリのある話し方を目指したものではない。また、これらの研究では聞き手とのインタラクティブなやり取りを想定していないため、聞き手が割り込んで発話するのに適した間（ま）の取り方については考慮されていない。

会話音声合成のためのデータ収録という観点では、様々な研究がある。それらは主に実際に会話を行わせて収録する方法と状況を設定して独話環境で収録する方法に分類される。前者の代表例としては、会話のシナリオを用意し、二人の声優に掛け合いをさせて収録したもの [Sugiura 14] がある。また、後者の代表的な方法としては、音声に込めたい感情（例えば「怒り」「喜び」

など) ごとに収録したもの [Adigwe 18, Burkhardt 05], 会話ロボットでの利用場面 (例えば「対話を盛り上げる」「相手を気遣う」など) ごとに収録を行ったもの [Iwata 11] などがある. 本研究は, 収録の容易性から状況を設定して独話環境で収録を行う後者の立場をとるが, 従来のような一発話ごとの音声収録ではなく複数の発話で構成される段落ごとに音声収録を行っている点, 重要な情報を印象付けるように意識させて発話させている点, 理解促進や質問誘発のための適度な間 (ま) を入れながら発話させている点など, 情報伝達を意識させた収録となっている点が異なる.

音声合成手法の観点では、DNN 音声合成モデルの入力として言語特徴の他に補助的な情報を与えることで合成音声を制御する試みが行われており、成果をあげている。例えば、入力 of 補助情報として、複数話者モデリングを実現するために話者コードを与える研究 [Hojo 16b, Hojo 18], 話者適応を実現するために i-vector や話者の性別を与える研究 [Wu 15], および両問題を単一モデルで表現するためにこれらの特徴量を組み合わせて与える研究 [Luong 17, Zhao 16], システム発話の意図を表現可能な音声合成を実現するために対話行為情報を与える研究 [北条 16a], 感情を表現可能な音声合成を実現するために感情情報を与える研究 [Inoue 17] などがある。本研究においてもこれらと同様に DNN に補助情報を加える立場をとるが、補助情報として発話の段落における位置や談話構造上の役割など、情報伝達に関連する段落単位での情報を用いている点が異なる。

6.3 システムの構成

本章では、開発する音声合成器の組み込み対象であるニュース伝達用の会話システムにおいて、ニュース記事が与えられてから、会話で発話するまでの流れを概説する（図 6.1）。

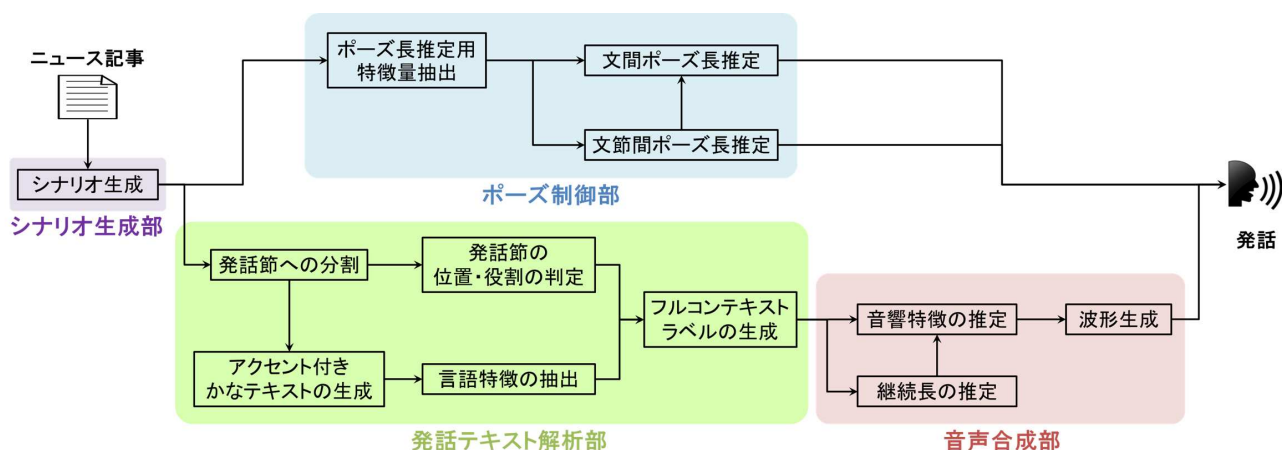


図 6.1: システム構成図

6.3.1 シナリオ生成部

シナリオ生成部では、ニュース記事から会話のシナリオを生成する。シナリオは主計画、副計画と呼ぶ、2種類の発話計画から構成される。主計画は、ニュース記事など、伝達対象の文書を要約し、口語化することで作られるもので、文書の要点を伝えることを目的としている。ユーザーが受け身で聴いている限り、システムは主計画に従った情報伝達を行う。一方、副計画は、会話進行の各時点におけるユーザーの反応予測に基づいて、それに対する応答を計画したもので、ユーザーの興味・理解度に応じた補足情報の提示を目的としている。ユーザーが能動的な情報要求を示してきた場合、システムは副計画に遷移し、情報の補完を行う。これらの発話計画を事前に準備することで、即応性の高い円滑な会話インタラクションを実現できる。

6.3.2 発話テキスト解析部

テキスト解析部では、シナリオ生成部で生成された発話文（各文に対して文圧縮と口語化処理を適用し生成された発話テキスト）から、音声合成モデルの入力となる言語特徴の抽出を行う。まず、発話文を「発話節」というより小さな発話単位に分割する。発話節は、一息で発話する単位であり、後述するように発話の理解しやすさを考慮し、新規情報が複数含まれないように作られる。次に、発話節ごとにアクセント句境界およびアクセント核の推定を行い、アクセントの位置やアクセント句の区切りを記したかなテキストを生成する。このかなテキストを入力として、音素の種類やアクセント句内のモーラ位置などの言語情報を抽出する。さらに、各発話節に対して元記事における段落内の位置や発話の役割を判定する。発話の役割として「核」「前置き」「補足」を定め、要約結果に基づいていずれかのラベルを各発話節に付与する。そして、発話節の位置・役割および言語特徴に関する情報を HTS³で用いられているラベル列（以下、フルコンテキストラベル）へ変換する。

6.3.3 音声合成部

音声合成部では、フルコンテキストラベルの情報から得られる言語特徴や発話節の位置・役割に関する特徴（以下、発話系列特徴）を入力として、ニューラルネットワークモデルを用いて継続長と音響特徴を推定し、推定された音響特徴から STRAIGHT ボコーダー [Kawahara 06] を用いて音声波形を生成する。

³<http://hts.sp.nitech.ac.jp/>

6.3.4 ポーズ制御部

ポーズ制御部では、発話文に対して形態素解析を行い、品詞や活用形に関する特徴量を抽出した後、発話文内の文節間のポーズ長と発話文間のポーズ長を回帰モデルで推定する。

6.4 音声データの収録と分析

ニュース記事をもとに人手で作成した発話原稿を女性声優に発話させ、音声データの収録を行った。6.4.1節では、音声合成用に収録した音声データとその分析結果について説明し、6.4.2節では、ポーズ長の推定用に収録した音声データとその分析結果について説明する。

6.4.1 音声合成用音声データの収録と分析

6.4.1.1 発話単位

話し言葉には、音声として発した情報が瞬時に消える揮発性と、発信した情報が瞬時に聞き手に受信される同時性という特徴がある [畠 87]。そのため、話し手は聞き手の負担を考慮して、一回の発話で伝える情報はなるべく少なく分かりやすい表現にすることが望まれる [Chafe 94]。

本研究では、伝えようとするニュース記事内の文を、述語や接続助詞などを手がかりに分割し、それぞれの区間に複数の新情報が含まれないように発話節を作成する。例えば、図 6.2 では連体修飾節と被修飾名詞を切り離し（主節化分割 [後藤 16]）、接続助詞を挟んだうえで、末尾を伝聞口調に書き換える（3章参照）。接続助詞を区切りとして分割された二つの文節列が発話節となり、一息で発話する単位となる。

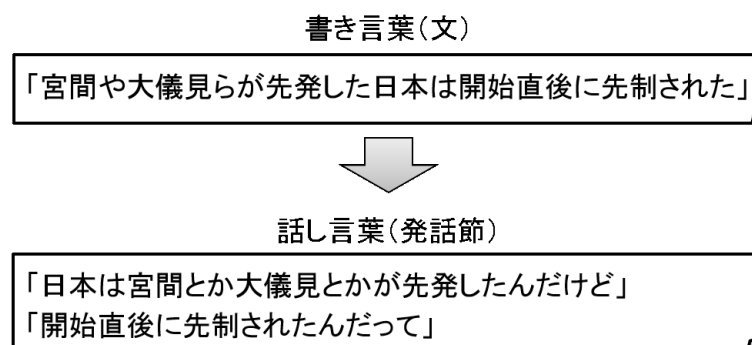


図 6.2: 発話節への分割例

6.4.1.2 発話節の役割

情報を伝える発話節の役割として「核」(Nucleus), 「前置き」(Front Satellite), 「補足」(Rear Satellite) の3種類を採用する. ここでの核は修辞構造理論 [Mann 88] における核 (Nucleus) と同等のもので, ニュースの重要な情報を伝える役割をもつ. 前置き, 補足は修辞構造理論における衛星 (Satellite) にあたるものである. 核の前に位置し, 情報を伝えるリード文の役割を果たす発話節を前置きとする. 核の後に位置し, 核の情報を補ったり, 核ほど重要ではないが付加すべき情報を伝える役割を果たす発話節を補足とする.

6.4.1.3 収録原稿の作成

収録原稿の例を図 6.3 に示す. 図 6.3 の例では各行が発話節となっており, (a) ~ (c) などの空行で区切られた単位が一つの段落を構成している.

段落	位置	役割	発話節の内容	
1	1	前置き	宇宙を観測しているロシアの電波望遠鏡があるんだけど	(a)
	2	核	それが正体不明の「強い信号」を検知したってって注目を集めてるんだって	(b)
	3	補足	世界中の天文学者に衝撃が走っているらしいよ	(c)
2	1	核	信号はヘラクレス座の近くの恒星から届いたんだって	(d)
	2	補足	その星は地球から約95 光年離れていて	(e)
	3	補足	惑星も持っているみたい	(f)
3	1	前置き	地球外文明の進捗度を表す基準っていうのがあって	(g)
	2	前置き	それと照らしあわせたんだけど	(h)
	3	核	地球よりもはるかに進歩した文明の可能性があるんだって	(i)
	4	補足	まだ確実なことはわからないんだけどね	(j)

図 6.3: 収録で用いた原稿の例

各発話節の役割は、図 6.3 の太字で示した (b), (d), (i) が核、核の前にある (a), (g), (h) が前置き、核の後ろにある (c), (e), (f), (j) が補足である。

原稿の作成にあたり、段落は 3 から 6 個程度の発話節からなり、必ず 1 個以上の核を持つように設計した。前置きと補足は複数存在する場合もあるが、核の位置によっては存在しないことを許す。段落の最初が核の場合は前置きが、最後が核の場合は補足が存在しなくなる。核としては、ニュース記事の見出しなどの重要な情報を含むものを選択した。また、収録にあたり、核となる発話節は太字で明記した。

合計 39 記事（114 段落、419 発話節）からなる原稿を作成した。図 6.4 は、それぞれの役割の発話節が段落内の何番目に現れているかを表したものである。この図から前置きは 1 番目や 2 番目といった段落の前方に、核は 2 番目や 3 番目に、補足は 3 番目以降に多く現れていることが分かる。

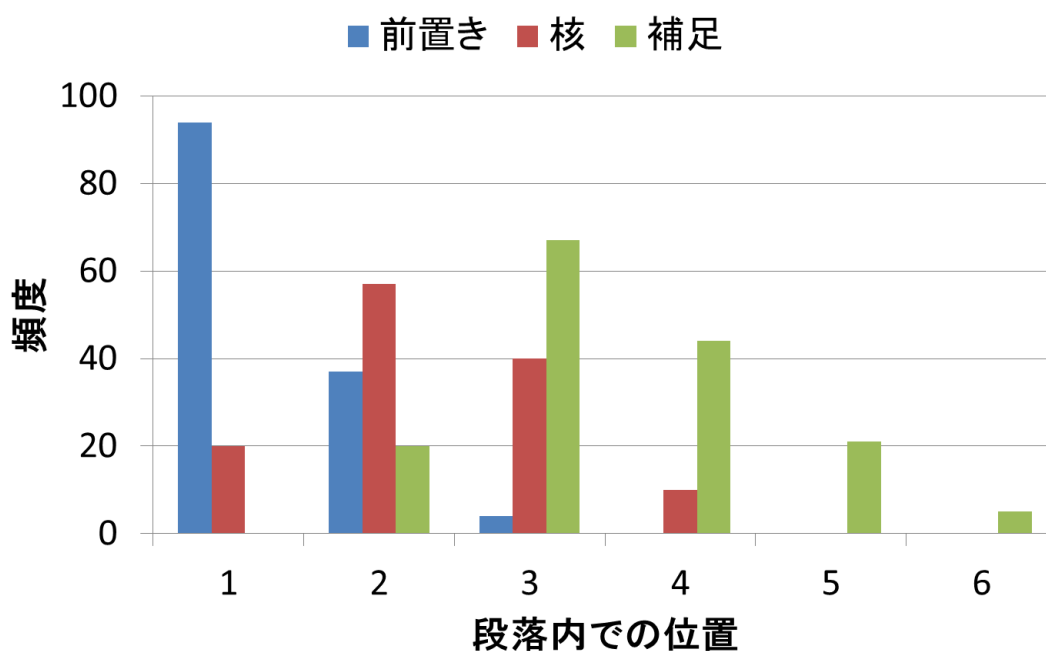


図 6.4: 役割ごとに見た発話節の段落内での出現位置の分布

6.4.1.4 収録概要

女性声優 1 人に原稿を発話させることで、段落ごとに音声データの収録を行った。収録時には以下の点を意識するように指示を与えた。

- (1) 太字で示された内容が相手にはっきりと伝わるように話すこと

- (2) 話し相手がうなずきながら聴いているものとして、途中で大きく間（ま）をあけることなく話すこと
- (3) 日常会話のような口調で、話し相手にニュースの内容を伝えるつもりで話すこと

6.4.1.5 発話節の位置と役割による韻律の違いに関する分析

収集した発話音声について、発話節の基本周波数（F0）や発話速度の変化に関する統計を調査した [福岡 17a, Fukuoka 17b]. 分析結果を付録 E 章に示す.

6.4.2 ポーズ長推定用音声データの収録と分析

音声合成用のデータの収録では、聞き手は話し手の話を傾聴しているものとして、相手のフィードバックを待つような大きな間（ま）をあけることなく話すように指示した. しかしながら、実際の会話では、相手に理解を促したり、相手の理解状態をうかがったりするために、適度に間（ま）を入れながら話を進める. 本節では、特にまとまった量の情報を伝える際の間（ま）の取り方に焦点を当てて音声収録を行った.

6.4.2.1 収録概要

テクノロジー系のニュース記事 100 個から人手で作成した発話原稿を女性声優 1 人に発話させ、その発話音声を収録した. ここでは、自然に出てくる間（ま）のデータを収集するために、原稿には核の情報や発話節の区切りを明示しなかった. 収録にあたり、話者に対して以下の点に注意するように指示した.

- (1) 話を聴いている相手に内容を伝える（理解させる）つもりになって発話すること
- (2) 重要な箇所は強調し、あまり重要でない箇所はさらっと伝えるなど、メリハリのある話し方を意識すること
- (3) 発話中でも割り込んで質問がくる可能性があり、質問がくると予想される箇所では不自然でない程度に間（ま）を入れること

6.4.2.2 データセットの分析

収録した音声から文節間のポーズ長および発話文間のポーズ長に関する時間情報を算出し、ポーズ長タグ付きコーパスを作成した。ここで、文節区切りには JUMAN++⁴と KNP⁵の結果を採用した。データセットの統計を表 6.1 に示す。また、発話文内の文節間のポーズ長と発話文間のポーズ長の分布をそれぞれ図 6.5 と図 6.6 に示す。

実際に収録音声から得られたポーズ長の例を表 6.2 に示す。この図から、発話文と発話文の間のポーズ長は 2 秒程度であり、発話文内の文節間のポーズ長は、連用節の後、つまり、発話節の区切りとなる場所で比較的長めのポーズが設けられていることが分かる。

表 6.1: ポーズ長タグ付きコーパスの統計

	文節間のポーズ長	発話文間のポーズ長
データ数	5327	507
最小値 (秒)	0.000	1.172
平均値 (秒)	0.210	2.657
最大値 (秒)	2.680	4.283

⁴<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

⁵<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

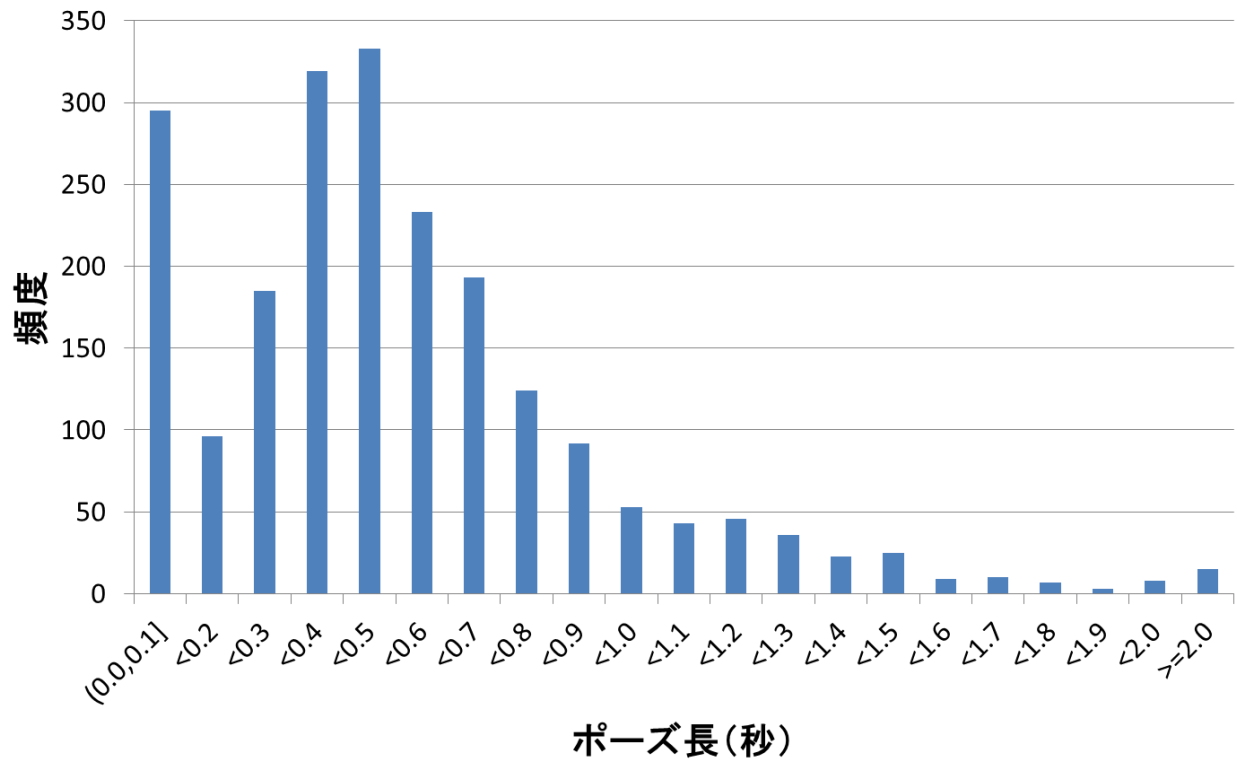


図 6.5: 発話文内の文節間のポーズ長（非ゼロ）の分布

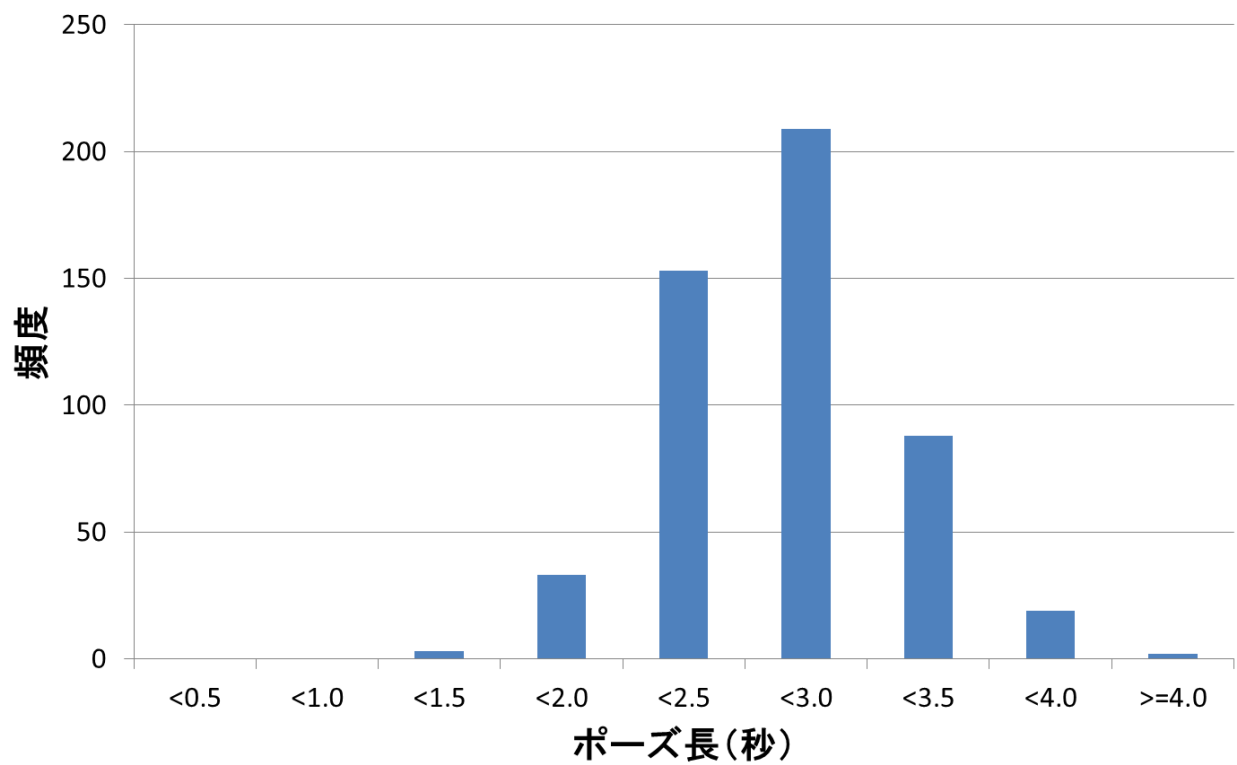


図 6.6: 発話文間のポーズ長の分布

表 6.2: 収録音声から得られたポーズ長の例（太線が発話文の区切り，細線が発話節の区切り）

発話節の役割	発話内容	ポーズ長（ミリ秒）
前置き	いま	181
	国内の	0
	自動車市場って	383
	ミニバンとか	0
	エコカーが	0
	主力なんだけど	650
核	高級スポーツカーの	0
	売れ行きも	0
	好調らしいよ	1620
前置き	日産が	0
	GT-R を	0
	9 年ぶりに	0
	改良して	716
前置き	今年の	0
	7 月に	0
	新モデルを	0
	発売したんだけど	853
核	996 万円以上も	0
	するのに	394
核	1 か月で	0
	年間販売計画台数の	129
	800 台を	0
	超える	0
	受注が	0
	あったんだって	2292

6.5 発話節の位置・役割を考慮した音声合成手法

前章では，ニュースを伝える発話系列音声において，発話節の位置と役割が韻律に影響を与えていることを確認した．本章では，従来の DNN 音声合成モデルにこれらを補助情報として組み込んだモデルを提案する．まず，音声合成器の入力となる言語特徴を得るための発話テキストの解析方法について説明する．次に，発話節の位置や役割（発話系列特徴）を考慮した DNN

音声合成モデルについて説明する．最後に，発話節間のポーズ長を推定するモデルについて説明する．

6.5.1 発話テキストの解析

6.5.1.1 発話節への分割

発話節への分割は，ニュース記事の各文に文圧縮・口語化処理を行うことで生成された発話を連用節や連体修飾節の位置で区切ることで行う（図 6.2）．ただし，連体修飾節の位置での分割において，被修飾名詞が形式名詞・副詞的名詞である場合や残りの文節数が2文節以下のように短い場合は分割しない．

6.5.1.2 発話節の位置・役割判定

発話節の位置は，段落内で何番目の発話節かを表す．発話節の役割は，要約処理で重要文として抽出された文を「核」とする．段落内で核の文よりも前にあるものを「前置き」とし，段落内で核よりも後ろにあるものを「補足」とする（図 6.3）．ただし，核の文でも複数の発話節に分割される場合は，最初のを前置きとし，以降のを核とする．また，核を持たない段落の文は，後ろに核を持つ段落が存在する場合は前置きとし，後ろに核を持つ段落が存在しない場合は補足とする．

6.5.1.3 アクセント付きかなテキストの生成

発話節のテキストを入力として，アクセント付きかなテキストを生成する．入力と出力の例を表 6.3 に示す．また，かなテキストの構成を表 6.4 に示す．アクセント位置およびアクセント句区切りは，LSTM を用いたモデルによって推定する [小野 18]．

表 6.3: アクセント付きかなテキストの生成例

入力	地球から 340 光年離れたところである惑星が見つかったんだけど
出力	ち) きゅーから／／さ@んびゃく／よんじゅっこ@ーねん／はな@れた／ところ@で／／ある／わく) せーが／みつ) かった@んだけど

表 6.4: アクセント付きかなテキストの構成

表記	意味
ひらがな	発話節内容（読み）
@	アクセント位置
/	アクセント句区切り 前の句が直後の句を修飾
//	アクセント句区切り 前の句が2つ以上後ろの句を修飾
:	ポーズ
)	母音の無声化 母音「い」と「う」が対象

6.5.1.4 フルコンテキストラベルの生成

フルコンテキストラベルは、音素ごとにコンテキストを考慮した属性（音素の種類やアクセント句内のモーラ位置など）を付与したラベル列である。本研究では、従来用いられてきた音素レベルから発話レベル（ここでは発話節レベル）の属性 [Iwata 11] に加え、発話節の位置や役割を付与したものを用いる。

6.5.2 発話系列特徴を考慮した DNN 音声合成モデル

従来の DNN 音声合成モデルに、発話節の位置や役割に関する特徴（発話系列特徴）を加えたモデルを提案する。

Zen らは、言語特徴を入力として音響特徴を予測する DNN 音声合成モデル（以下、音響モデル）を提案した [Zen 13]。Zen らのモデルでは、音響モデルの入力とする継続長には自然音声から得られた継続長が用いられているが、Wu らの DNN 音声合成ツールキット Merlin⁶ [Ronanki 16, Wu 16] では、継続長の推定にも DNN（以下、継続長モデル）を用いている。本研究では、モデルの作成、学習に Merlin を使用した。

提案モデルの概要を図 6.7 に示す。まず、フルコンテキストラベルから DNN の入力となる言語特徴を抽出する。言語特徴には、音韻に関する質問（子音かどうかなど）の解答といったバイナリ特徴や、発話文のアクセント句の数といった数値特徴が含まれる。継続長モデルでは、言語特徴を入力として各音素の継続長を出力するニューラルネットワークを構築する。出力された継続長をもとにフレーム特徴 [Wu 16] を計算し、言語特徴とともに音響モデルの入力とする。

提案モデルでは、入力に言語特徴やフレーム特徴に加えて発話節の位置・役割を表す特徴（発

⁶<https://github.com/CSTR-Edinburgh/merlin>

話系列特徴)を与える．発話節の位置に関する情報は，現在の発話節が段落内の何番目であるかを1次元の整数値で表現する．発話節の役割に関する情報は，対象の発話節が前置き，核，補足のどの役割に該当するかを one-hot な3次元のベクトルで表現する．つまり，該当する役割の次元が1になりその他が0の値をとる．これら4次元の発話系列特徴を言語特徴やフレーム特徴に加えて，継続長モデルと音響モデルの入力に与える．

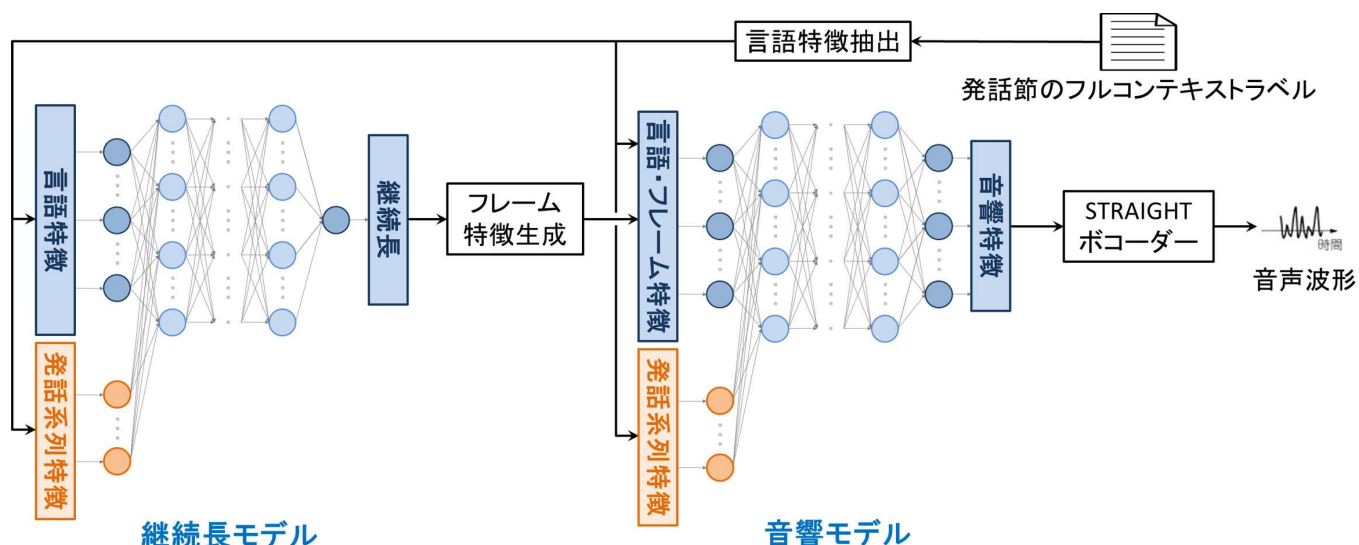


図 6.7: 発話系列特徴を考慮した DNN 音声合成モデル

6.5.3 ポーズ長の推定

6.4.2 節の分析で発話文内の文節間のポーズ長と発話文間のポーズ長の分布が大きく異なっていることが分かった．そこで，発話文内の文節間のポーズ長を推定するモデル（以下，文節間ポーズ長推定モデル）と発話文間のポーズ長を推定するモデル（以下，文間ポーズ長推定モデル）の2つのモデルを用いてポーズ長を推定する．なお，音声合成時にも発話節内のポーズ長が推定されるため，音声合成結果と間（ま）の制御を統合して会話を行う際は，発話節内のポーズ長は音声合成結果のものを採用し，本節のモデルで推定されたポーズ長は発話節間の間（ま）の部分に用いる．

6.5.3.1 文節間ポーズ長推定モデル

発話文内の文節間のポーズ長を推定するモデルを提案する．モデルには順方向1層，逆方向1層の双方向LSTMを用いる．文節間のポーズ長は前後の文節のLSTMの状態に基づいて推定する（図6.8）．入力は文節ごとに文字の系列を1層のLSTMでリカレントに埋め込んだものを与える．この際，文字の one-hot 表現に加え，単語の品詞や活用形の情報を与える．そして，実際に声優が発話したときのポーズ長を予測するように誤差逆伝播法でモデルの学習を行う．

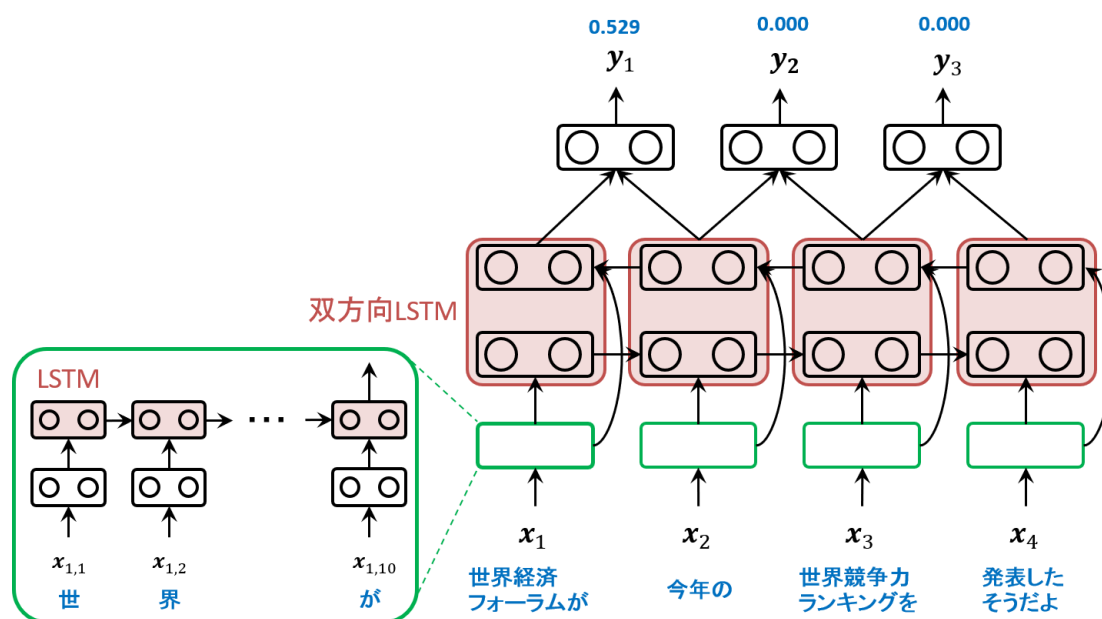


図 6.8: 文節間ポーズ長推定モデル

6.5.3.2 文間ポーズ長推定モデル

発話文間のポーズ長を推定するモデルを提案する．発話文間の間（ま）の役割として，発話内容の咀嚼・理解や質問誘発の他に，話題変更の合図が考えられる．そのため，現在の発話内容だけでなく，次の発話内容も考慮してポーズ長を決定する必要がある．ここでは，6.5.3.1 節で学習したモデルを用いて，現在の発話文における順方向 LSTM の最終文節の隠れ層の値と次の発話文における逆方向 LSTM の先頭文節の隠れ層の値を特徴量として，リッジ回帰により発話文間のポーズ長を推定する（図6.9）．また，特徴量には前後の発話文の単語数も加える．

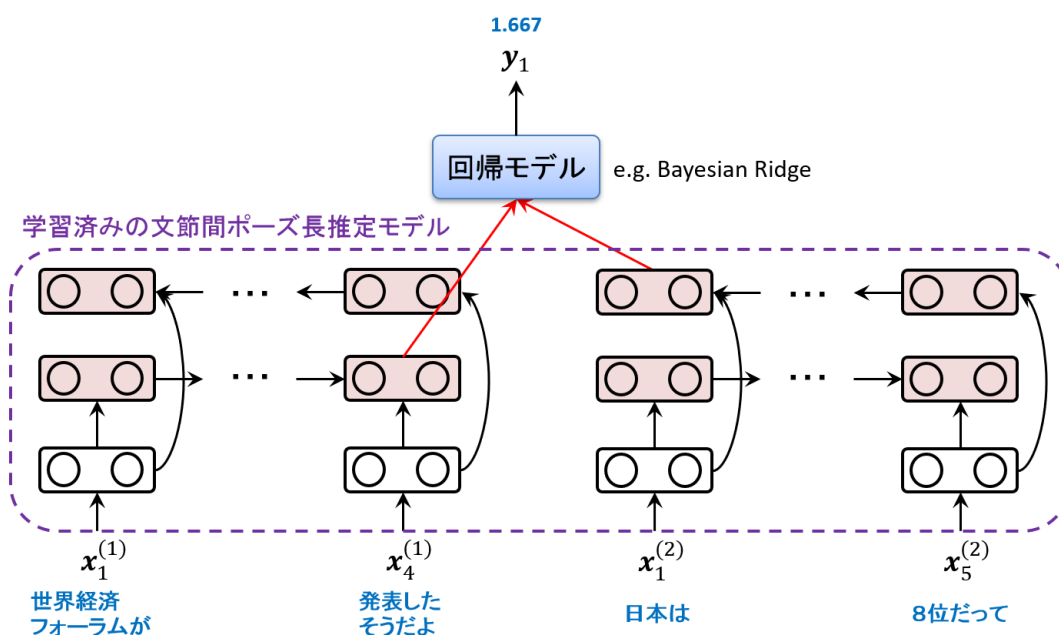


図 6.9: 文間ポーズ長推定モデル

6.6 評価実験 1: モデルの性能評価

収録した音声データを用いて、音声合成モデルとポーズ長推定モデルの評価を行った。

6.6.1 音声合成

発話系列特徴を加えないモデルをベースライン，発話系列特徴を加えたモデルを提案モデルとして性能の比較を行った。

6.6.1.1 実験設定

データセットとして 6.4.1 節で収録した 419 個の発話節からなる音声データとシナリオを用いた。ここで、発話節の位置・役割の付与，および，アクセント付きかなテキストの作成までの処理は人手で行った。

合成フィルタには STRAIGHT [Kawahara 06] を使用した。音響特徴量は、0-39 次元のメルケプストラム，1 次元の対数 F0，5 次元の帯域ごとの非周期性指標および，これらの Δ ， Δ^2 と，1 次元の有声／無声区間判定（有声区間なら 1，無声区間なら 0）の計 139 次元からなる。音声

データからの音響特徴量の抽出は付録E章と同様の条件で行った。

DNNの隠れ層は6層でユニット数は1024に設定し、中間層の活性化関数にはtanh、最適化アルゴリズムにはAdamを使用した。

6.6.1.2 実験結果

10分割交差検定でモデルの出力と元データの音響特徴量の誤差を計算した結果を表6.5に示す。ここで、F0-RMSE, dur-RMSEは基本周波数、音素継続長の平均二乗誤差平方根（Root Mean Square Error）を表し、MCD, BADはメルケプストラム歪み（Mel-Cepstrum Distortion）、帯域ごとの非周期性指標の歪み（Band-Aperiodicity Distortion）を表している。なお、音響モデルの入力には、原音声の継続長を使用した。

基本周波数、継続長に関して、4次元の発話系列特徴を加えることでベースラインに比べて誤差が小さくなることが確認できた。また、メルケプストラム、非周期性指標についても発話系列特徴を加えることでわずかに改善が見られた。

表 6.5: 発話系列特徴の有無による性能の比較

	F0-RMSE	dur-RMSE	MCD	BAD
ベースライン	36.420	4.779	5.064	0.660
提案モデル	35.435	4.768	5.047	0.659

6.6.2 ポーズ長の推定

6.4.2節で作成したポーズ長タグ付きコーパスを使用し、文節間ポーズ長推定モデルと文間ポーズ長推定モデルに関して、有効な特徴量の検討および性能の評価を行った。

6.6.2.1 文節間のポーズ長の推定

モデルの構造として、文節を構成する単語や文字のBag-of-Words (BoW)を入力として与えたとき、これらをLSTMでリカレントに埋め込んだときの平均二乗誤差平方根（RMSE）の比較を行った。さらに、単語や文字の基本素性の他にJUMAN++やKNPを適用して得られる

品詞や活用形、活用型、カテゴリ、ドメイン、固有表現の他、『基本語データベース』⁷の単語親密度および『日本語教育語彙表 ver1.0』⁸の語彙難易度を追加素性として使用し、有効な素性の組み合わせについて検討した。なお、ポーズ長の推定結果が読点「、」の与え方に左右されないように、入力文から「、」を取り除いた。実験では、埋め込み層および隠れ層の次元を 100 次元に設定し、活性化関数には tanh、最適化アルゴリズムには Adam を使用した。

10 分割交差検定により評価した結果を表 6.6 に示す。実験の結果から、文節の文字をリカレントに入力したときの結果が総じて良く、追加素性として品詞と活用形および活用型の情報を与えたときに最も RMSE が小さな値を示した。さらに、推定値と正解の誤差のヒストグラム（図 6.10）を確認したところ、0.2 秒未満の誤差の割合が 77%であった。

表 6.6: 文節間ポーズ長推定モデルの結果（RMSE）

追加素性	単語 BoW	文字 BoW	単語 RNN	文字 RNN
	0.252	0.260	0.251	0.233
品詞	0.234	0.240	0.231	0.224
活用形	0.239	0.242	0.230	0.230
活用型	0.245	0.255	0.237	0.229
カテゴリ	0.253	0.261	0.249	0.236
ドメイン	0.252	0.259	0.248	0.233
固有表現	0.250	0.260	0.250	0.236
親密度	0.263	0.264	0.250	0.243
難易度	0.260	0.262	0.249	0.242
品詞+活用形+活用型				0.219

⁷https://hon.gakken.jp/reference/special/jiten/kihongo_db/index.html

⁸<http://jhlee.sakura.ne.jp/JEV.html>

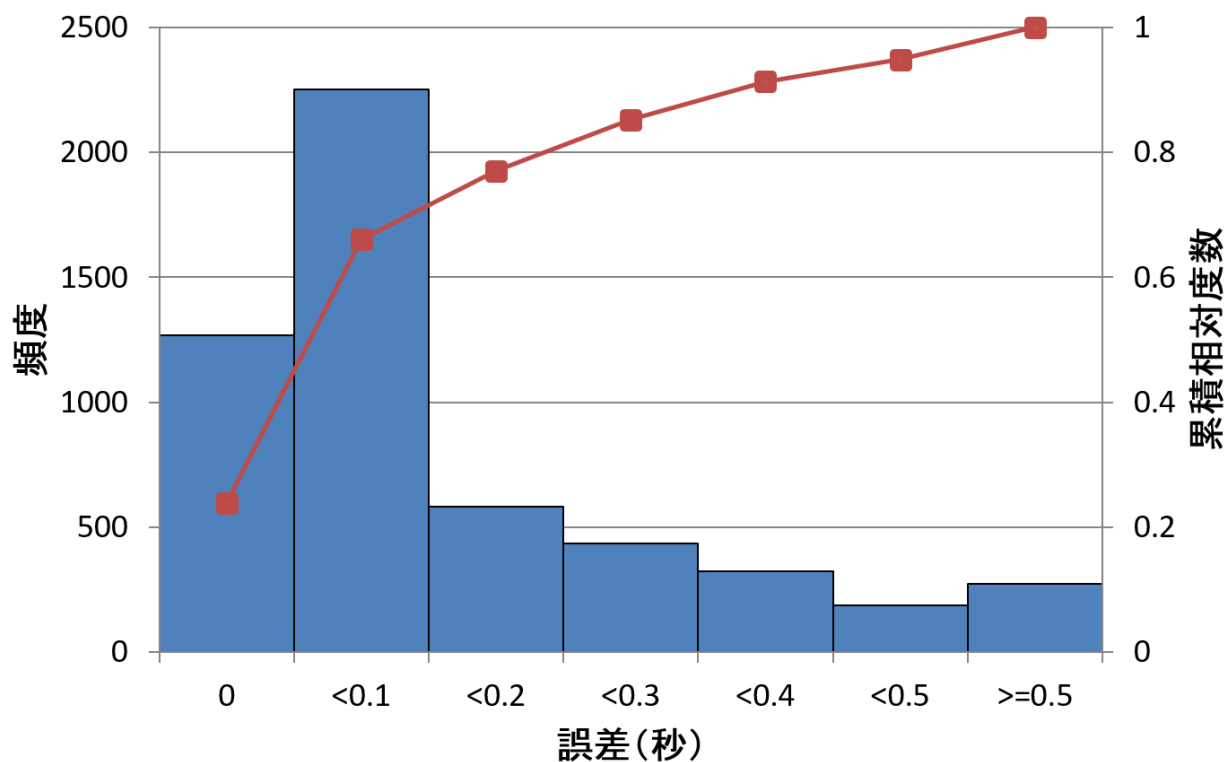


図 6.10: 文節間ポーズ長の誤差（折れ線は累積相対度数）

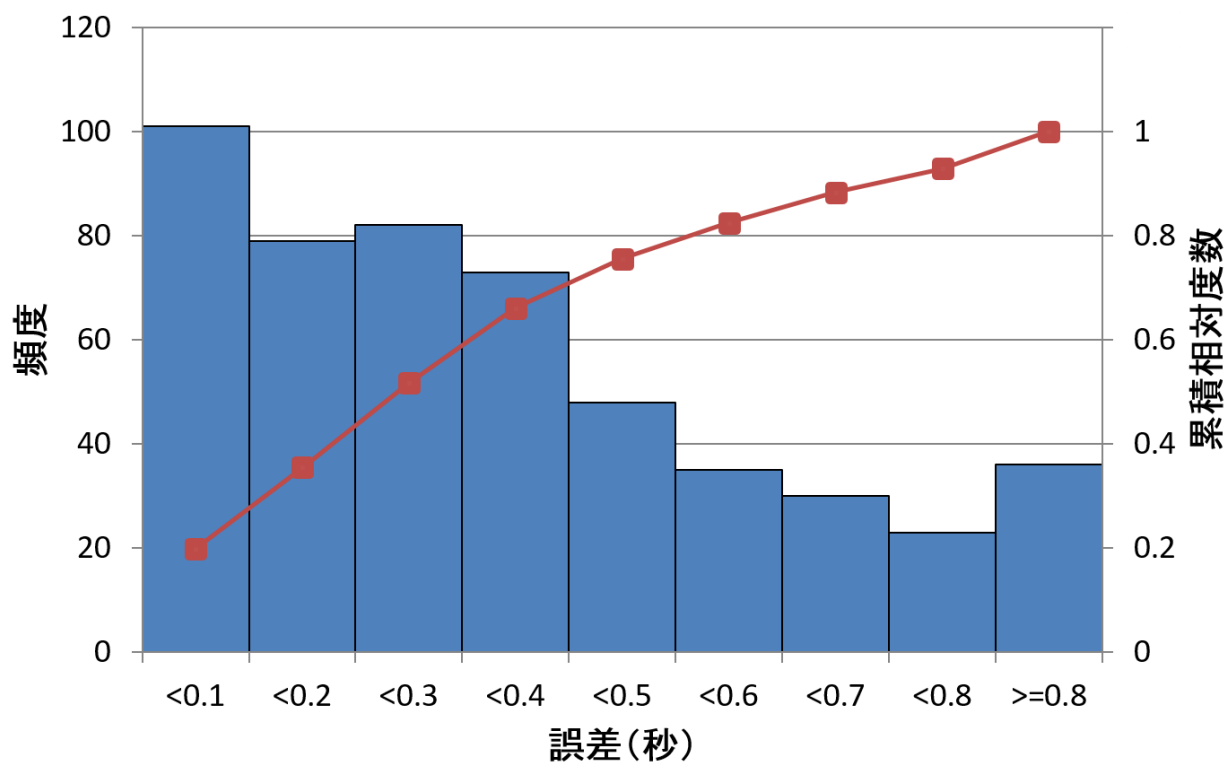


図 6.11: 文間ポーズ長の誤差（折れ線は累積相対度数）

6.6.2.2 発話文間のポーズ長の推定

特徴量として前節で学習した文節間ポーズ長推定モデルの LSTM の隠れ層の値を用いた。予備実験の結果、現在の発話文における順方向 LSTM の最終文節の隠れ層の値と次の発話文における逆方向 LSTM の先頭文節の隠れ層の値の組み合わせが最も良い結果を示した。また、scikit-learn⁹の回帰モデルを一通り比較した結果、BayesianRidge モデルが最も良い結果を示した。

10 分割交差検定で評価したところ、RMSE は 0.450 であった。ポーズ長が前後の発話の長さに影響されるという研究報告もあり [鴨井 11]、前後の発話の文字数、単語数、文節数を素性に加えて比較したところ、RMSE は文字数を加えた場合 0.440、単語数を加えた場合 0.436、文節数を加えた場合 0.437 を示した。この結果から発話の長さに関与するこれらの特徴量を加えることの有効性が確認された。さらに、単語数を加えたモデルにおいて、推定値と正解の誤差のヒストグラム（図 6.11）を確認したところ、0.5 秒未満の誤差の割合が 75% であった。

6.6.2.3 ポーズ長の推定結果の例

10 分割交差検定で得られたポーズ長の推定結果の例を表 6.7 に示す。この結果から、提案モデルの推定結果が人の話し方に近い間（ま）の取り方になっていることが分かる。また、音声収録において指示した「質問が予測される箇所で不自然でない程度に間（ま）を入れる」という部分に関して、実際に声優は 1 発話文目の「あるんだけど」の後、および、3 発話文目の「使ってて」の後に十分な間（ま）をあけており、システムもそれを再現できている。

⁹<http://scikit-learn.org/stable/>

表 6.7: ポーズ長の推定結果（ミリ秒）（「推定結果の間」と「人の間」のいずれかが100 ミリ秒以上であった箇所．太線は発話文の区切り，細線は発話節の区切り）

発話内容	推定結果の間	人の間
ジュノー っていう	61	210
無人探査機があるんだけど	1085	1311
それが撮影した	329	417
木星の北極と南極の画像を	412	581
NASA が公開したそうだよ	2700	2630
太陽系の	214	0
他の惑星には見られない	459	641
雲とかの気象が見れるみたい	2503	2553
撮影には	308	375
オーロラマッピング装置 を使ってて	832	1272
これまで観測されなかった	394	376
周囲と温度が違うスポットを捉えたんだって	2521	2453
それで	544	667
木星の	238	262
オーロラの謎が解明できるんじゃないかって	511	388
期待されてるそうだよ		

6.7 評価実験2: 情報を伝える話し方としてのふさわしさに関する主観評価

音声合成モデルの学習に用いる音声データの発話単位，間（ま）の制御の有無，発話系列特徴の利用の有無の3つの条件を変えて合成音を作成し，情報を伝える話し方としてのふさわしさについて評価を行った．

6.7.1 実験概要

音声合成モデルの学習に用いる音声データの発話単位に関して，「単発話」と「段落発話」の比較を行った．ここで，単発話は，学習データを一文ずつ発話・収録することを表す．一方，段落発話は，6.4.1 節で説明した，核を明示し，複数の発話節からなる発話系列を段落単位で発話・

収録することを表す．段落発話モデルの学習には，6.4.1 節で収録した音声データを使用し，単発話モデルの学習には，音素バランスのとれた会話調の 700 発話 [Iwata 11, 小林 14] を段落発話と同じ声優で別途収録した音声データを使用した．

間（ま）の制御に関しては，発話節間のポーズ長として，6.5.3 節で提案したポーズ長推定モデルの推定結果を用いたものと，6.4.1 節で収録した音声データにおける核から核へ移行する際の平均時間に相当する 0.5 秒に固定したものとを比較した．

発話系列特徴の利用に関しては，発話節の位置・役割に相当する発話系列特徴を DNN の補助情報として利用した場合と利用しない場合を比較した．

10 個のニューストピックについて，これらの組み合わせによって表 6.8 に示す 4 タイプの音声合成し，下記 3 つのタスクについて一対比較による評価を行った．サンプル音声は，1 記事あたり 3 段落程度の発話系列からなる約 1 分の音声である．

表 6.8: 比較する合成音声の種類

合成音声 タイプ	単発話 モデル	段落発話 モデル	間（ま） の調整	発話系列 特徴
1	✓			
2	✓		✓	
3		✓	✓	
4		✓	✓	✓

タスク A 間（ま）の調整の有無に関する比較（タイプ 1 とタイプ 2 の合成音声の比較）

タスク B 単発話モデルと段落発話モデルの比較（タイプ 2 とタイプ 3 の合成音声の比較）

タスク C 発話系列特徴の有無に関する比較（タイプ 3 とタイプ 4 の合成音声の比較）

評価は「情報を伝える話し方としてどちらがよりふさわしいと感じたか」という観点で，被験者 15 人に対して実施した．各タスクは被験者ごとにランダムな順序で実施し，比較する音声に関しても提示順序をランダムにして聴かせた．また，音声を複数回聴くことを許容した．サンプル音声は，1 記事あたり 3 段落程度の発話系列からなる約 1 分の音声である．なお，図 6.1 のシナリオ生成部で主計画を作成する処理（要約と口語化）は人手で行った．

6.7.2 実験結果

各タスクの実験結果を図 6.12, 図 6.13, 図 6.14 に示す.

間（ま）の調整の有無に関する比較では, 一定の短い間（ま）で淡々と発話する話し方よりも, 内容に応じて間（ま）の長さを調節する話しの方が情報を伝える話し方としてふさわしいとする結果が得られた.

単発話モデルと段落発話モデルの比較では, 段落発話モデルの方が良い結果が得られた. 段落発話モデルの学習データは, まとまった量の情報を伝えることを強く意識して収録した音声データである. そのため, 一発話に限ってみても, 抑揚の付け方や話速が単発話モデルよりもダイナミックに変化しており, 情報を伝える話し方としてよりふさわしいと判断されたためだと考えられる.

発話系列特徴の有無に関する比較では, 発話系列特徴を加えたときの方が良い結果が得られた. 発話系列特徴を加えることによって, 発話系列全体として核が際立つようなメリハリのある話し方を実現できたためだと考えられる.

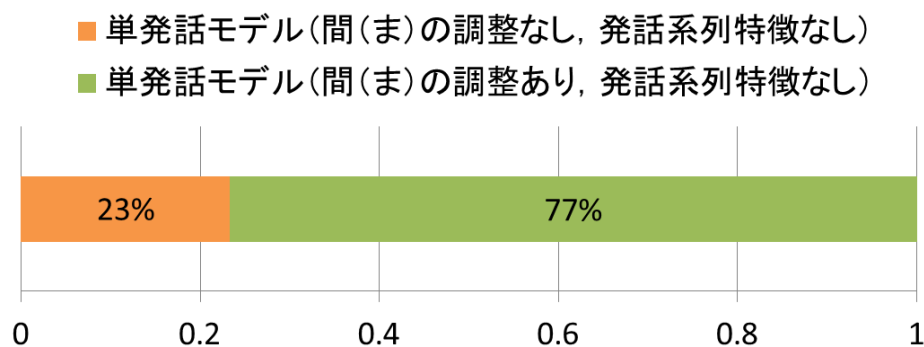


図 6.12: 【実験 2】タスク A : 間（ま）の調整の有無に関する比較

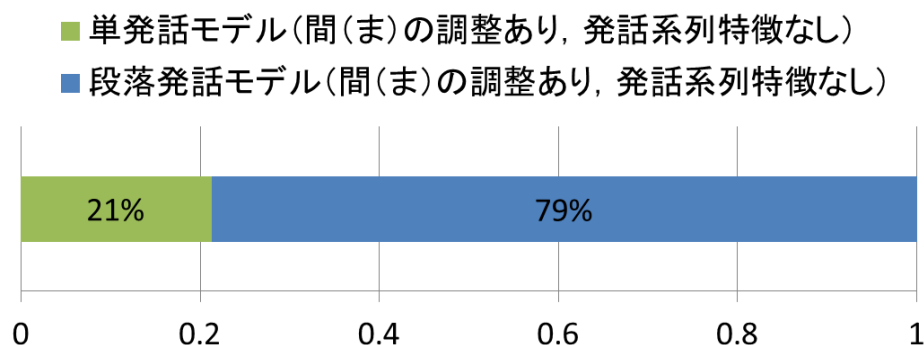


図 6.13: 【実験 2】タスク B : 単発話モデルと段落発話モデルの比較

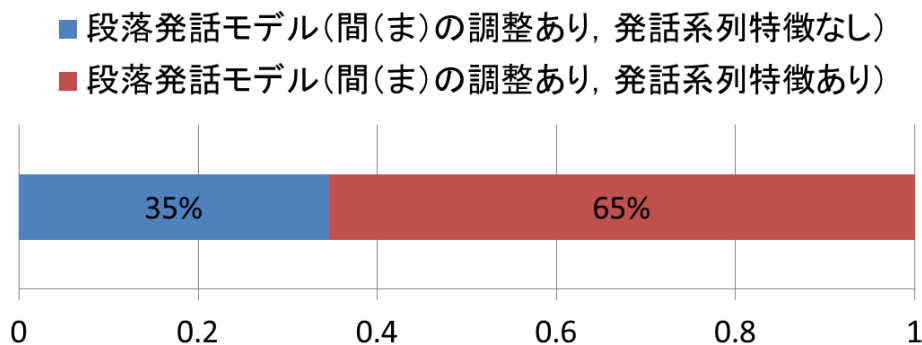


図 6.14: 【実験 2】 タスク C: 発話系列特徴の有無に関する比較

6.8 評価実験 3: 理解度に及ぼす影響の評価

6.7 で説明した 3 つの操作（間（ま）の調整，段落発話データの使用，発話系列特徴の追加）を加えることで，ニュースの理解度にどの程度影響を及ぼすか評価を行った．

6.8.1 実験概要

10 個のニューストピックについて，6.7 節で説明した 4 種類の音声（表 6.8）を合成し，6.7 節で述べた A，B，C の 3 つのタスクについて評価を行った．

各タスク 10 人ずつ合計 30 人に対して実験を行った．各タスクの評価では，被験者ごとにタイプ i ($i \in \{1, 2, 3\}$) の音声 が 5 トピック，タイプ $i+1$ の音声 が 5 トピックとなるように配分し，ランダムな順序で実施した．1 トピックは 3 段落から 5 段落の発話系列から構成される．

実験では，被験者に対して 1 トピック音声 が聴き終わるごとに質問に回答させた．質問は 1 トピックにつき核の内容に関する質問と核以外の内容に関する質問を 1 問ずつ用意した．例えば，図 6.3 のようなニュースに対し，核の内容に関する質問として「信号を発していた恒星は何座の近くにありますか」，核以外の内容に関する質問として「信号を観測したのはどこの国ですか」のような質問を作成した．質問はなるべくニュースの前半の内容から作成した．回答は記述式で，完全に正解している場合を 1 点，部分的に正解している場合を 0.5 点として採点を行った．例えば，フルネームを答えるべきところを名字のみが正解していた場合などは部分点を与えた．被験者には「1 つのトピックで音声を聴いてよいのは一度だけ」「メモを取ってはならない」「分からない場合は未記入でよい」ことを指示した．

6.8.2 実験結果

各タスクについて正解率を算出した結果を表 6.9, 表 6.10, 表 6.11 に示す.

タスク A の結果 (表 6.9) から, 短く単調な間（ま）で話すよりも適度な長さの間（ま）を入れて話す方が内容が頭に入りやすいことが分かった.

タスク B の結果 (表 6.10) から, 従来の 1 文単位の読み上げ音声を用いて学習したモデルよりも, 相手に伝えることを強く意識させたうえで段落単位で発話させた音声を用いて学習したモデルの方が頭に入りやすいことが分かった.

タスク C においては, 発話系列特徴の利用によって, それを利用しない場合よりも正解率が向上することを期待したが, ほとんど差が見られなかった (表 6.11). また, 発話系列特徴なしの合成音声では核の内容に関する質問の正解率と核以外の内容に関する質問の正解率の差が小さく, 発話系列特徴ありの合成音声ではこの差が大きくなることも期待したが, 有意な差は見られなかった.

6.7 節の実験結果 (図 6.14) も踏まえると, 発話系列特徴を利用することで, 情報を伝える話し方に関する主観的な評価は改善するものの, 今回のような実験設定では, 理解度に関する客観的な評価までは改善できないことが分かった.

タイプ 1 の結果を 1 としたときの改善率を図 6.15 に示す. タイプ 1 の音声に対して間（ま）の調整を行うことで 1.13 倍理解度が向上し, さらに段落発話モデルを用いることで 1.17 倍理解度が向上することが分かった.

表 6.9: 【実験 3】タスク A : 間（ま）の調整の有無に関する比較

	間（ま）の調整なし	間（ま）の調整あり
正解率（全体）	0.595	0.675

表 6.10: 【実験 3】タスク B : 単発話モデルと段落発話モデルの比較

	単発話モデル	段落発話モデル
正解率（全体）	0.525	0.615

表 6.11: 【実験 3】タスク C : 発話系列特徴の有無に関する比較

	発話系列特徴なし	発話系列特徴あり
正解率（核）	0.770	0.780
正解率（核以外）	0.660	0.640
正解率（全体）	0.715	0.710

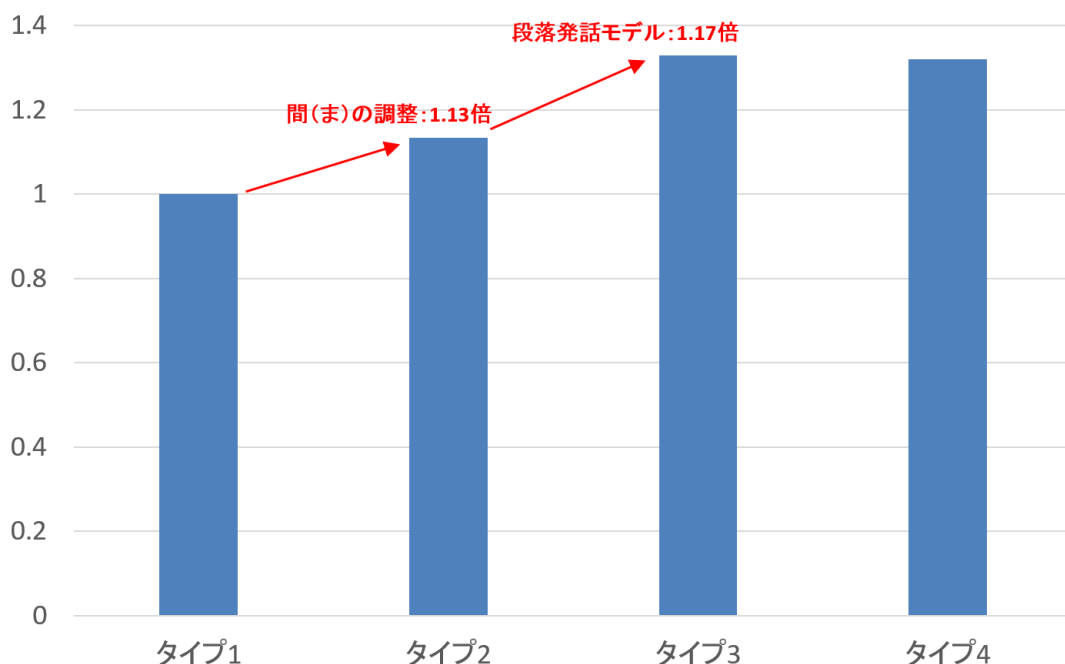


図 6.15: 【実験 3】 タイプ 1 の結果を 1 としたときの理解度の改善率

6.9 評価実験 4: 質問しやすさに関する評価

6.7 節の評価は、受け身で聴いたときの評価であった。そのため、特に間（ま）の評価において、ただ聴くだけにしては長い間（ま）も存在しており、間（ま）の有効性が十分に評価できていない。

間（ま）の役割には咀嚼・理解の促進の他に、フィードバックの誘発がある。インタラクティブなやり取りを通して情報を伝える会話システムにおいて、ユーザーがフィードバックを示しやすい間（ま）の取り方になっていることが重要である。そこで、本節では、ユーザーが示すフィードバックの中でも特に情報伝達効率（EoIT; Efficiency of Information Transfer）に寄与する質問を対象に、間（ま）の調整の有無が質問のしやすさにどのような影響を及ぼすか評価を行った。

6.9.1 実験概要

音声合成モデルには発話系列特徴を加えて学習した段落発話モデルを使用した。このモデルによって生成された合成音声に対して提案手法によって間（ま）の調整を行ったシナリオと、6.4.1 節で収録した音声データにおける核から核へ移行する際の平均時間に基づいて発話節間の

ポーズ長を 0.5 秒に固定した，間（ま）の調整なしのシナリオをそれぞれ 2 トピックずつ合計 4 トピック用意し，被験者 16 人に対して会話させた．

4 つのニューストピックには，アルファベット数文字からなる未知の単語が含まれている．未知語はニュース記事中の略語（e.g. “HMX”）の文字の間にアルファベットを挿入することで作成した（e.g. “HvMpX”）．評価実験で利用したシナリオの例を表 6.12 に示す．図 6.1 のシナリオ生成部で主計画を作成する処理（要約と口語化）は人手で行った．表 6.12 の “HvMpX” が未知語であり，未知語を含む発話節の終了地点が質問がくると予想される箇所である（表 6.12 の「…ぶつけて」の後）．表 6.12 のポーズ長は提案手法で推定した結果であり，間（ま）の調整なしの場合はすべて 0.5 秒とした．なお，4 つのトピックの質問想定箇所における推定結果のポーズ長はすべて 0.5 秒よりも長い．

被験者には，知らない単語があった際には「何それ」と質問するように指示した．1 トピックにつき間（ま）の調整ありの会話が 8 人，間（ま）の調整なしの会話が 8 人となるように配分し，被験者ごとに 4 つのトピックをランダムな順序で実施した．ただし，間（ま）の調整ありの 2 トピックと間（ま）の調整なしの 2 トピックが前半と後半に分かれるような順番で実施した．実験後のアンケートでは，前半 2 つの会話と後半 2 つの会話でどちらが質問しやすかったかについて回答させた．

表 6.12: 質問しやすさの評価実験で用いたシナリオの例（太線は発話文の区切り，細線は発話節の区切り）

発話節の内容	ポーズ長（秒）
はやぶさ 2 を搭載したロケットの打ち上げが成功したらしいよ	1.95
はやぶさ 2 のミッションはある小惑星からサンプルを採取してくることなんだって	2.13
惑星に到着したら	0.14
赤外線カメラとかを使って	0.35
表面の状態を観測したり	0.48
HvMpX を惑星にぶつけて	0.82
その衝撃でできたクレーターを観測したりするんだって	2.37
到着は 2018 年で	0.24
18 ヶ月間のミッションを経て	0.53
2020 年にサンプルを地球に持ち帰るんだって	

6.9.2 実験結果

実験後のアンケートで得られた質問しやすさに関する評価結果を図6.16に示す。この結果から、約7割の被験者は、内容に応じて間（ま）の長さを調節する話し方が、一定の短い間（ま）で淡々と発話する話し方よりも質問しやすいと評価したことが分かる。

未知語を含む発話節の発話終了時刻を基準として何秒後に質問があったか、間（ま）の調整なしの会話時と間（ま）の調整ありの会話時のそれぞれについて箱ひげ図で表したものを図6.17に示す。また、質問するまでの平均時間を算出した結果を表6.13に示す。ここで、質問想定箇所よりも前で質問が来た場合の時刻は負の値としている。また、「え、何それ」や「ん、何それ」のような質問は「え」や「ん」のタイミングを質問のタイミングとした。これらの結果から、間（ま）の調整を行ったときの方が質問するタイミングが平均的に遅いことが分かる。これは、間（ま）の調整なしの会話の方が話すテンポが早いため、システムのテンポに引きずられてユーザーの質問するタイミングが早くなったためだと考えられる。

質問を条件ごとに分類したときの事例数を図6.18に示す。未知語について質問した事例数は間（ま）の調整なしの方が多かったものの、0.5秒以上後で質問した事例数は間（ま）の調整ありの方が多かった。間（ま）の調整なしの会話は単調なテンポで発話が行われるため、システムが次に発話するタイミングを予測しやすい。今回はその間（ま）が0.5秒であったため、間（ま）の調整なしの会話では0秒以上0.5秒未満で質問したものの数が多かったと考えられる。一方、間（ま）の調整ありの会話は内容に応じて間（ま）の長さが変化する。中には0.5秒よりも短いものもあるが、発話文の区切りや質問が予想される箇所では適度に長い間（ま）が挿入される。このようなメリハリのある話し方をすることで、被験者はシステムのテンポに惑わされずに自分のタイミングで質問でき、その結果、少し間（ま）をおいたタイミングで質問が行われたと考えられる。

図6.16の結果も踏まえると、内容に応じて間（ま）の調節を行う話し方が、短い単調な間（ま）で発話する話し方よりも質問しやすい間（ま）を実現できたと考えられる。

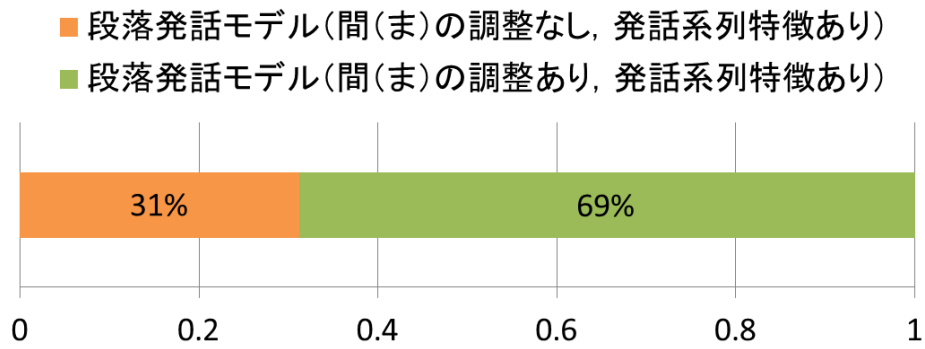


図 6.16: 質問しやすさに関する評価結果

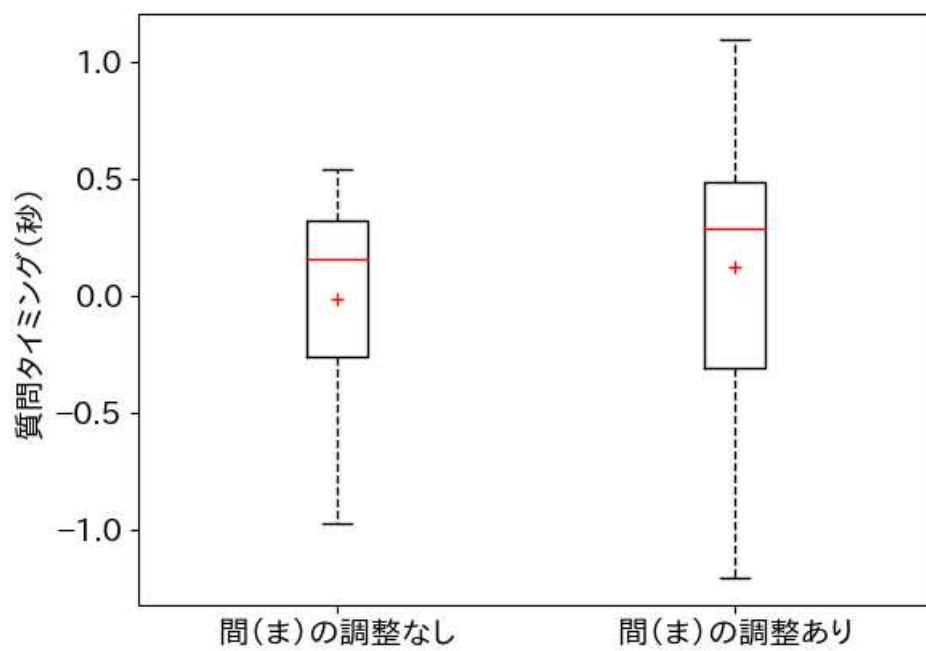


図 6.17: 質問想定箇所を基準としたときの質問タイミングの分布

表 6.13: 質問想定箇所を基準としたときの質問するまでの平均時間 (秒)

間（ま）の調整なし	間（ま）の調整あり
-0.016	0.123

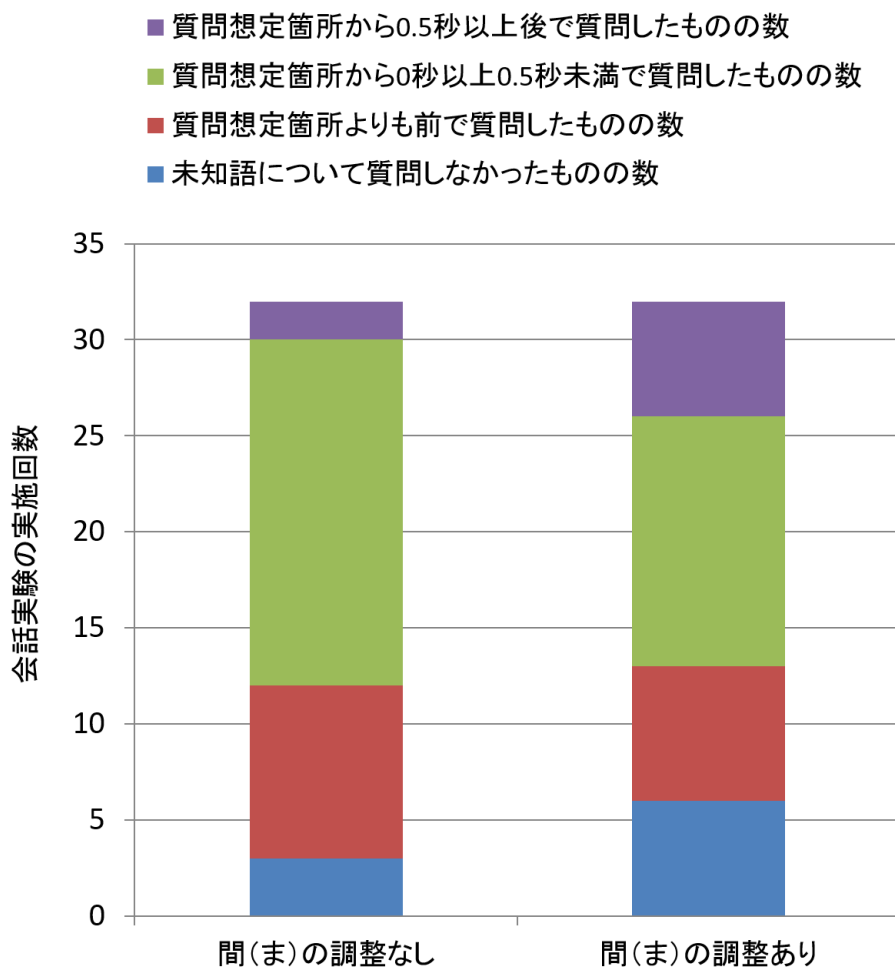


図 6.18: 条件ごとに見た質問の事例数

6.10 質問を挟みやすくするための間(ま)の調整に関する予備実験

実際に人がシステムの発話中に割り込んで質問する際、質問するまでにどれくらいの時間を必要とするかについて調査する実験を行った。

6.10.1 実験設定

難しい単語を含む重文の発話を対象に「何それ」と質問させ、質問に要する時間を計測した。例えば、表 6.19 のようなシステム発話があったとき、被験者は「アフアール猿人」について「に含まれるんだけど」の後に、「何それ」と質問する。実験は 11 人の被験者に対して行い、合計 16 個の質問対象語（e.g. アフアール猿人）を含む 14 個のニューストピックについて会話させた。ここで、実験の順序効果をなくすためにトピックの順番は被験者ごとにランダムに入れ替

えて実施した。また、実験では、なるべく実際の会話と利用状況を似せるために、被験者の相槌や感想などの発話も許容するとともに、上記質問に対する回答も提示した。

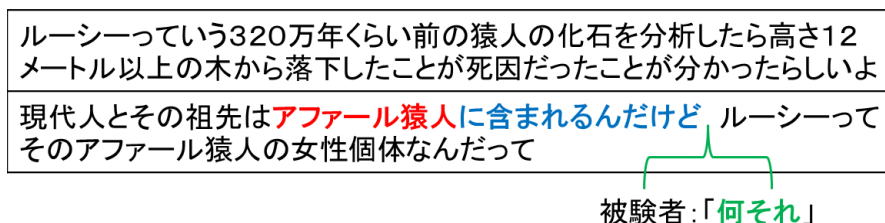


図 6.19: 質問に必要な間（ま）の調査実験における質問の例

6.10.2 実験結果

質問までの平均時間が早かった順に被験者を A から K まで番号付けし、質問するまでにかかった時間を被験者ごとに箱ひげ図で表したものを図 6.20 に示す。この結果から質問に必要な時間は被験者ごとに大きく異なることが分かる。次に、質問までの平均時間が早かった順に質問対象語を a から p まで番号付けし、質問するまでにかかった時間を質問対象語ごとに箱ひげ図で表したものを図 6.21 に示す。この結果から文脈依存性はそれほど大きくないことが読み取れる。次に、ポーズ時間と質問カバー率の関係を図 6.22 に示す。この結果から 0.7 秒のポーズを入れることで 80% の質問をカバーできることが分かる。しかしながら、被験者ごとに確認してみると、A, B, C, E のように 0.7 秒のポーズで 100% の質問をカバーできる被験者も入れば、K のように 44% しかカバーできない被験者もいる。

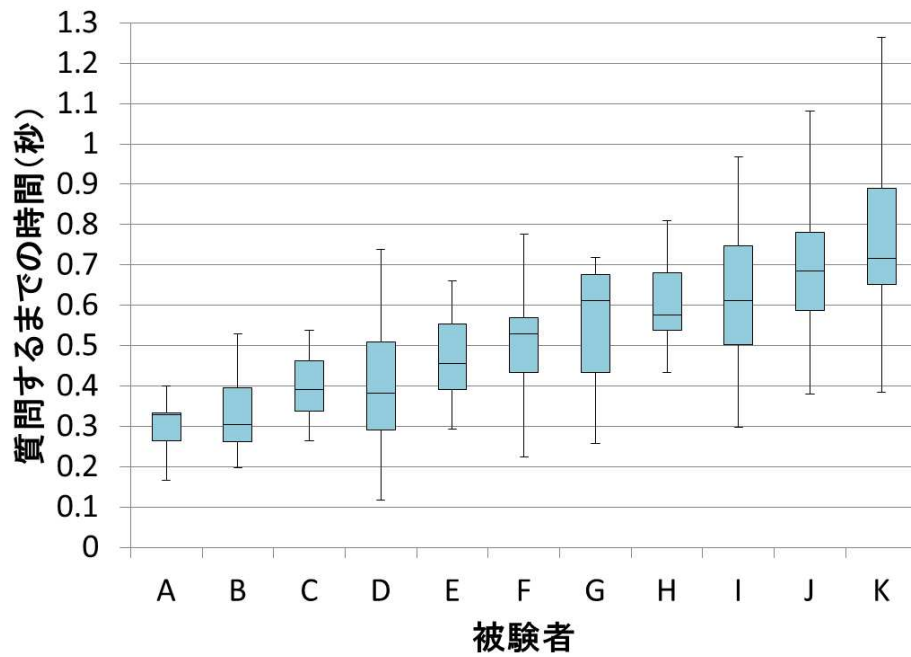


図 6.20: 質問に要する時間（被験者ごと）

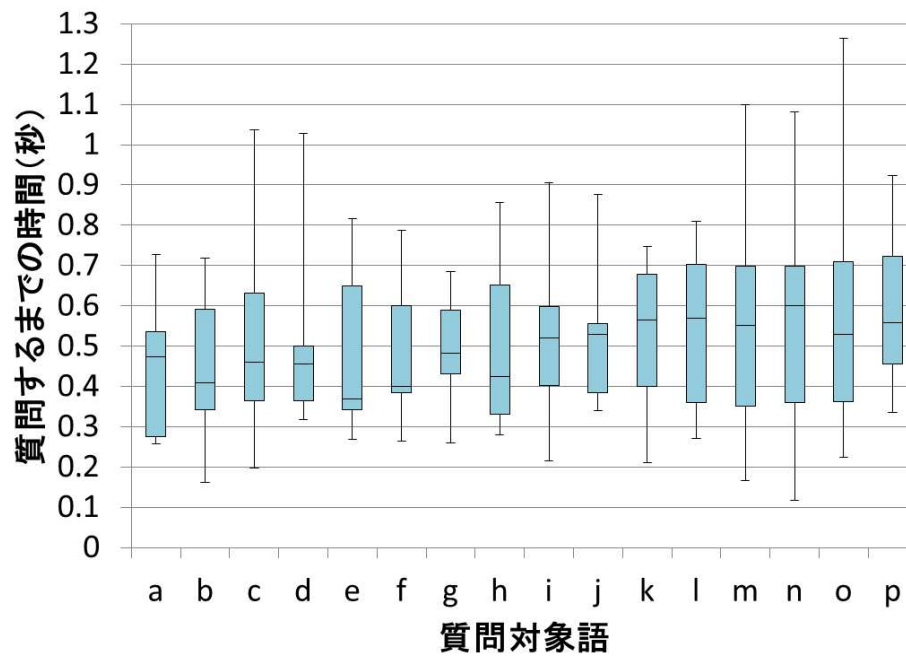


図 6.21: 質問に要する時間（質問対象語ごと）

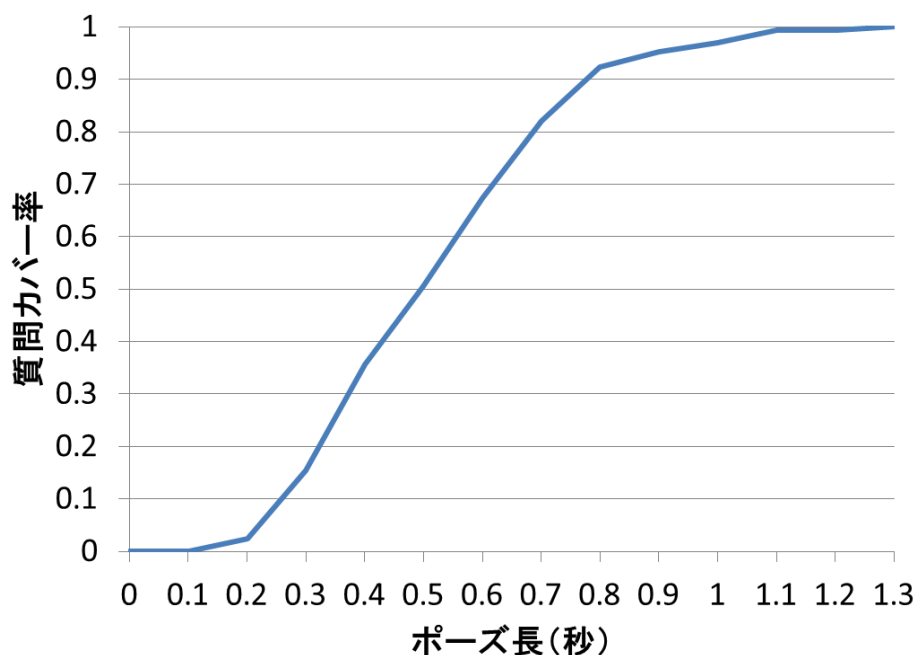


図 6.22: ポーズ長と質問カバー率の関係

6.11 発話スタイルの転移学習に関する検討

本研究で収録した音声は、感情的な分類で言えば、ニュートラルな話し方である。しかしながら、当然、人間が情報を伝える場合、感情が常に一定ということはなく、会話の状況・伝える内容によって発話スタイルが変化する。例えば、スポーツの優勝のニュースであれば、興奮気味に話すのに対し、殺人事件のニュースであれば、不快感がにじみ出る。このような伝達内容に応じた感情の切り替えを可能とすべく、現在所持している音声データから感情に関する発話スタイルの転移学習がどの程度可能か検討した。

6.11.1 実験設定

発話スタイルの分類を図 6.23 に示す。これは、岩田らに対話状況に応じた音声の使い分けを想定して分類したものである [Iwata 11]。所有している音声データには、話者 X のものとして {N, S, E, R, A} の単発話データと N の段落発話データ、話者 Y のものとして N の単発話データと N の段落発話データがある。このとき、5 種類の発話スタイルで話者 B の段落発話音声を合成することを目指す。

話者 ID、感情 ID を発話系列特徴に加えて補助情報として、提案モデル (図 6.7) に与える。補助情報の与え方を表 6.14 に示す。この区別に従って音声データを混ぜ、モデルの学習を行った。

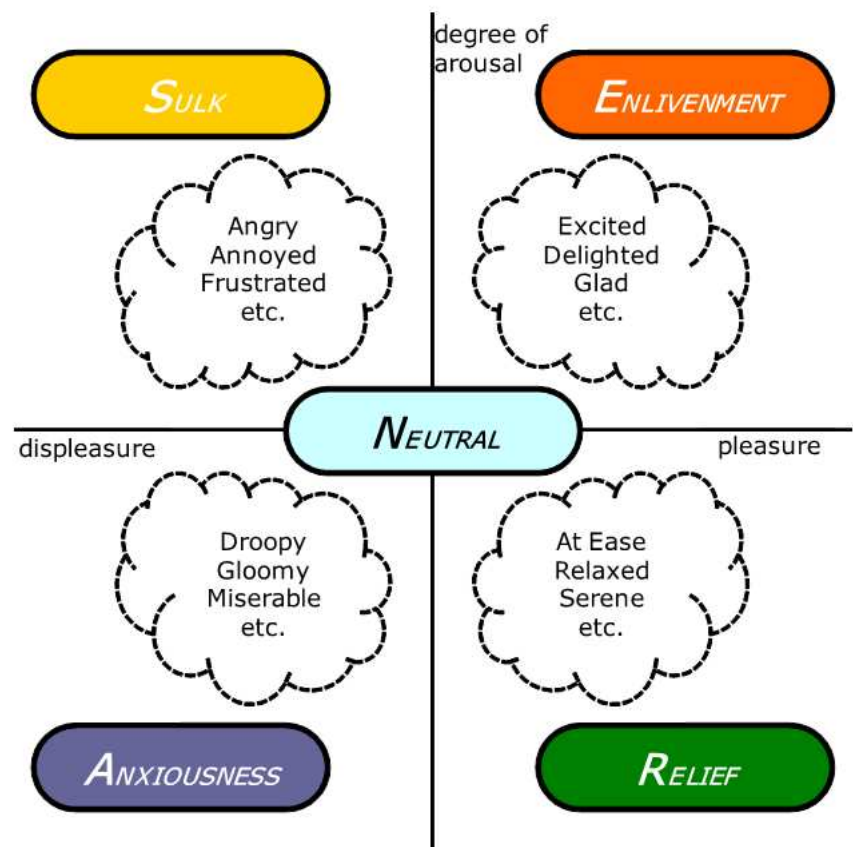


図 6.23: 発話スタイルの分類 [Iwata 11]

表 6.14: 発話スタイルを転移学習させるための補助情報の与え方

位置	役割			話者		感情				
	前置き	核	補足	X	Y	N	S	E	R	A
整数	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1

6.11.2 実験結果

発話系列特徴と感情を指定し、話者 Y で合成した音声を確認したところ、感情の込め方がスタイルごとに変化していることが確認できた。しかしながら、抑揚の付け方が単発話データの話し方に引きずられ、メリハリが弱まってしまった。今後は、感情に関するスタイルのみを転移させ、メリハリが失われないように学習する方法を検討する。

6.12 まとめと今後の課題

ニュースのようなまとまった量の情報を伝える会話システムにおいて、談話構造上の発話の役割に応じたメリハリのある話し方を可能とする音声合成システムを開発した。

書き言葉と話し言葉の違いに留意して、ニュース記事をもとに発話原稿を作成し、音声収録を行い、発話の位置や役割を考慮した DNN 音声合成モデルと、文節間・発話文間のポーズ長を推定するモデルを提案した。

提案した間（ま）の推定モデルによって、人の作る間（ま）に近い適切な間（ま）を推定でき、挿入すべき間（ま）の長さを一定の短い間（ま）で淡々と発話する話し方よりも、聴きやすく頭に入りやすく、さらに質問しやすい話し方を実現できることが確認できた。今後は、会話の活性度に応じて間（ま）の長さを変化させるなど、会話の状況やユーザーとの親密度に応じて、間（ま）を制御する方法について検討する。

従来の1文単位の読み上げ音声で学習したモデルよりも、情報伝達を意識させ段落単位で発話させた音声を用いて学習したモデルの方が、情報伝達のための発話としてふさわしいと感じるとともに、内容理解の面においても優れていることが分かった。今後は、学習データの量を増やすことで、音素バランスを整えるなど学習データの質を改善することで、これらの評価の成績が向上するかどうかについて検討する。

発話の段落における位置や談話構造上の役割を補助情報に用いて合成することにより、情報伝達にふさわしい聴きやすい話し方になることが確認できた。しかしながら、その効果が、理解度を向上させるところまでには至らなかった。今回の実験条件では、被験者は集中して音声を聴くことができたため、多少メリハリの弱い話し方であっても十分に内容を理解することができたと考えられる。今後は、運転しながら、料理しながらといった多重タスクの条件下で、理解度テストを実施し、求められるメリハリの水準を上げたときの発話系列特徴の導入効果について検討する。

今回の研究では、システムが主体的に情報伝達を行うという状況に焦点を当てて検討を行った。そのため、ユーザーはうなずきながらシステムの話聞いていることを仮定し、声優が演技して発話した音声を分析対象とした。しかしながら、より活性化された会話において、システム発話のリズムは、ユーザー発話のリズムに影響を受けることが予想される。これらの相互作用をモデル化する立場に立つならば、より統制の少ない会話環境で収集した音声の分析を進めることも重要である。

第7章 統合システムの評価

7.1 情報伝達効率に関する評価

主計画と副計画からなるシナリオに沿って、自由なタイミングで受動／能動のモード遷移が可能な提案システムの評価を行った。評価では、会話を通じてユーザーがどれだけ欲しい情報を引き出せたかに関する主観評価を実施した。また、モードの切り替えに制約がある従来型の会話システムと比較して、情報行動にどのような違いが見られるかに関する調査も行った。

7.1.1 実験設定

提案する会話システムと従来型の会話システムの比較評価を行った。以下、自由なタイミングでのモード遷移が許される提案システムを用いた会話タスクを「自由対話タスク」と呼び、受動モードにより一定の情報獲得を行った後に一問一答形式の会話を行う従来型の会話システムを用いた会話タスクを「一問一答タスク」と呼ぶ。

ここで従来型の会話システムとしては、最初にニュースのサマリを読み上げてから、QAモードへ移行する吉野らのシステム [吉野 14, Yoshino 15a, Yoshino 15b] を模したシステムを使用した。一問一答タスクでは、まず、システムがニュース記事を要約した内容を伝える。この間、利用者はしゃべらずに、その内容を聴く。その後、質問や感想があれば、システムに話しかける。さらに、ユーザーが興味がありそうな場合には、システムは主体的に情報を付加する。

自由対話と一問一答のいずれにおいても、話すことがなくなった時点で、被験者が「会話終了」と発話することでタスクが終了する。また、本システムの利用方針として、受け身で聴くことも許容しているため、「無理して話す必要はない」ことを事前に伝えた。

自由対話と一問一答のタスクは、それぞれ合計5トピック用意し、1トピックずつ交互に評価を実施した。各タスクの評価で使用したニュース記事のタイトルと文数を表 7.1, 表 7.2 に示す。なお、要約モデルとして Yahoo コーパスで学習したモデルを使用し、主計画を構成する重要文の数 K は5に設定し、主計画の、発話から発話へ遷移する際は2秒に設定した。

実験では、まず、被験者はタスクを確認し、そのタスクに従ってシステムと会話を行う。次に、この会話で使用したニュース記事を読み、各文に対してどれだけその情報が欲しかったかを以下の5つの項目から選択し、アンケート用紙に記入する。

5. 伝えてほしい
4. どちらかと言えば、伝えてほしい
3. どちらでもよい
2. どちらかと言えば、いらない
1. いらない

それが終わったら、次のトピックに移り、同様に会話と評価を順番に行う。実験は11人の学生に対して実施した。

以降の説明では、両タスクのアンケートの結果、以下の興味度の数値が高かった順に被験者をAからKと呼称する。

$$\text{興味度} = \sum_{i=1}^5 i \times (i \text{ 点の数}) \quad (7.1)$$

表 7.1: 自由対話タスクの評価実験で使用したニュース記事のタイトルと文数

記事タイトル	文数
IoT 端末使い外出先で救急対応、セコムが新サービス	22
ドコモがキッザニアに「ロボット研究開発センター」	21
理研など、リアルタイム降水予報を開始 30 秒更新	17
欧州発、電気自動車シフト 「脱石油」世界の潮流に	32
海外視聴者取り込め フジテレビ、AI で自動翻訳	46

表 7.2: 一問一答タスクの評価実験で使用したニュース記事のタイトルと文数

記事タイトル	文数
レアアース価格上昇 中国で環境規制強化	19
「健康になる家・掃除不要な家」募集 国交省補助事業	26
「2050 年に太陽光 200GW」、JPEA が長期ビジョン発表	15
野菜の産地偽装、通信用レーザーで判定 NTT グループ	35
ロボット産業、愛知の第3の柱に	43

7.1.2 実験 1: 自由対話タスクの実験結果

5 点の文に関して、主計画に基づく発話で伝えられた数、ユーザーからの働きかけによるインタラクションによって副計画から伝えられた数、会話で伝えられなかった数を積み上げたグラフを図 7.1 に示す。同様に 4 点以上の文の内容をどれだけ伝えられたかを示したグラフを図

7.2に示す．次に，1点の文に関して，主計画に基づく発話で伝えてしまった数，ユーザーとのインタラクションによって副計画から伝えてしまった数，正しく省けた数を積み上げたグラフを図7.3に示す．同様に2点以下の文の内容をどれだけ省くことができたかを示したグラフを図7.4に示す．

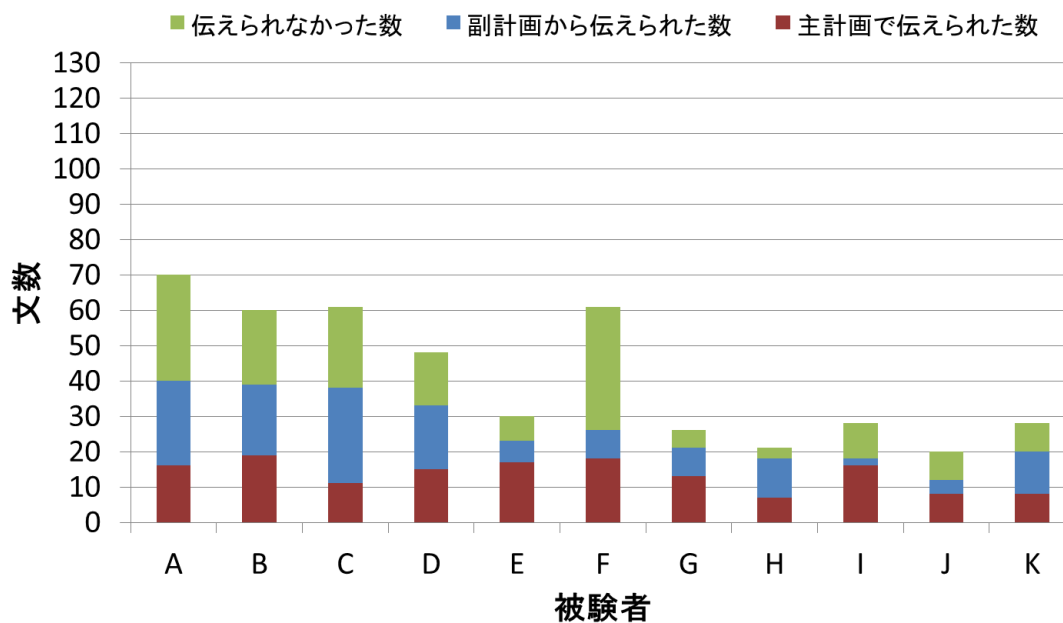


図 7.1: ユーザーが強く興味を持つ文（5点をつけた文）を伝えられたかどうかに関する分析結果（自由対話）

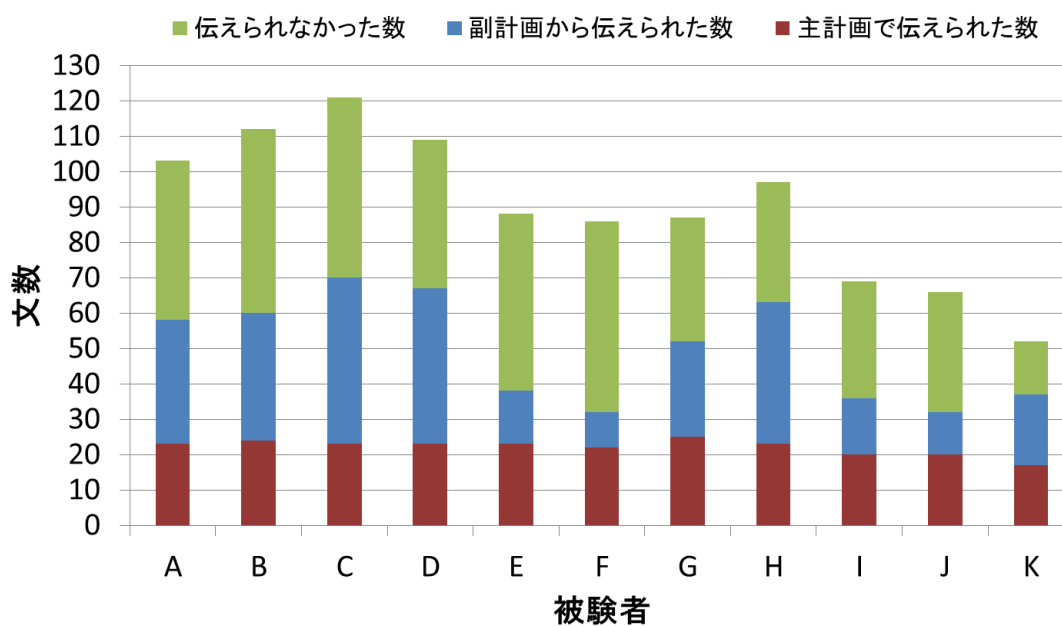


図 7.2: ユーザーが興味を持つ文（4点以上をつけた文）を伝えられたかどうかに関する分析結果（自由対話）

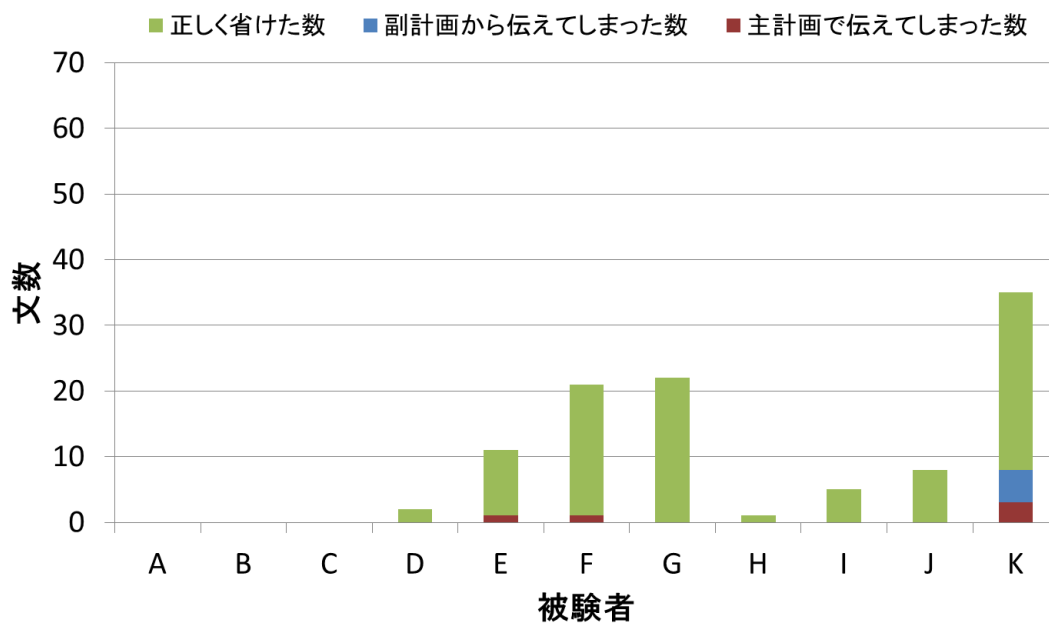


図 7.3: ユーザーが全く興味を持たない文（1点をつけた文）を除外できたかどうかに関する分析結果（自由対話）

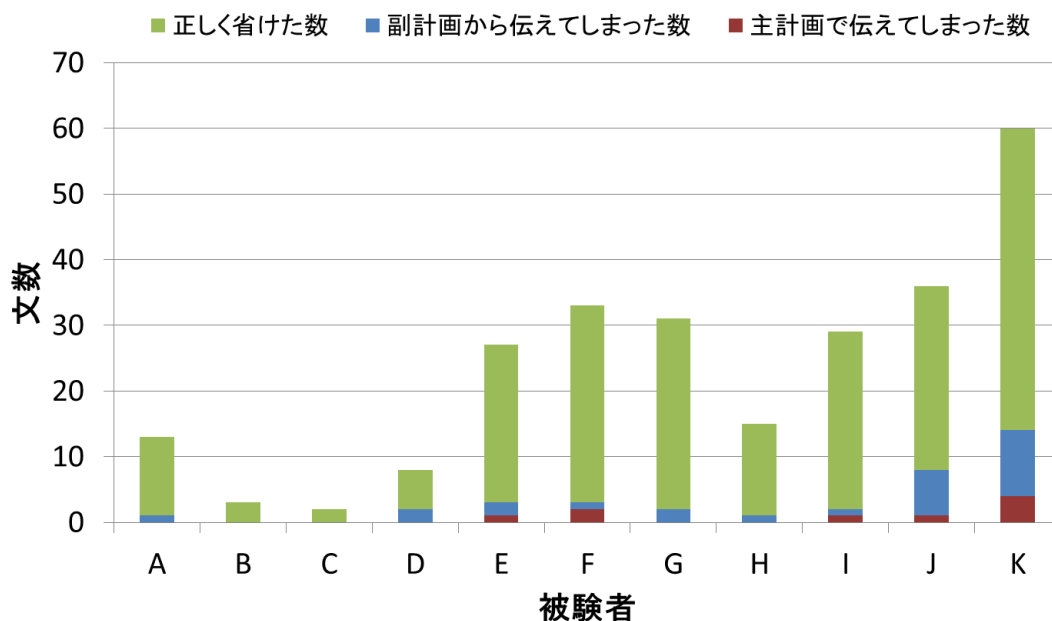


図 7.4: ユーザーが興味を持たない文（2点以下をつけた文）を除外できたかどうかに関する分析結果（自由対話）

ユーザー発話の種類を「質問・確認」(e.g. 誰が開発したの), 「感想」(e.g. 面白そうだね), 「納得」(e.g. なるほど, 確かに), 「相槌」(e.g. うん, ふーん), 「その他」に分類し, その出現回数を図 7.5 に示す. ここで, 「その他」には, 「感謝」や「笑い」, 「驚き」, 「自問」といった発話の他, 頻度は少ないが重要な現象として「待機要求」(e.g. それってさー) や「反復要求」(e.g. もう一回言って) が含まれる. システムは「待機要求」を受け付けた場合には数秒間停止し, 「反

復要求」を受け付けた場合にはもう一度発話を繰り返す。

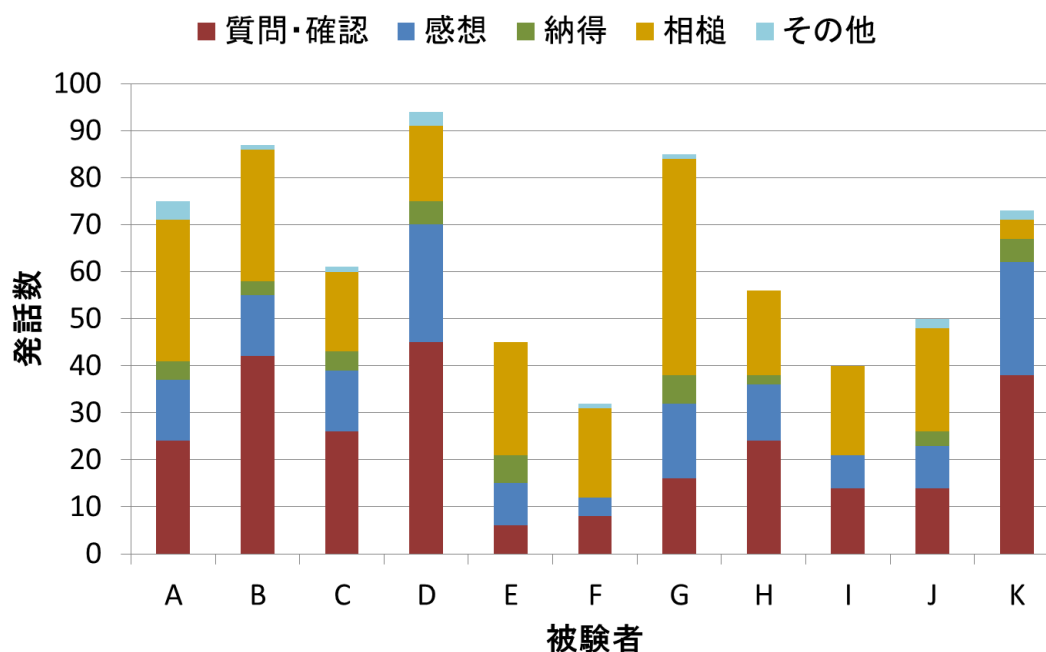


図 7.5: ユーザー発話の種類と回数（自由対話）

まず、A から D の被験者の結果に着目すると、他の被験者に比べて、トピック全体に対する関心が高く、欲しい情報が多いことが見て取れる。どれだけ欲しい情報を引き出せたかのグラフ（図 7.1, 図 7.2）を観察すると、元々、主計画で提示できる欲しい情報の割合は 2,3 割程度であったが、会話によって欲しい情報を 6 割近く引き出せていることが分かる。次に、被験者 F の結果に着目すると、トピックに対する関心が高いにもかかわらず、あまり発話していないため、会話で引き出せた欲しい情報の割合が低いことが分かる。このように、ニュースに興味はあるが、あまり話したくない（受け身で聞きたい）ような利用者に対しては、主計画を大きめに設計するなどの方策が有効だと考えられる。次に G から K の被験者の結果を観察すると、比較的高い割合で 5 点の文をカバーできていることが分かる。一方で、知らない情報をどれだけ省けたかのグラフ（図 7.3, 図 7.4）を観察すると、他に比べて知らない情報が多いことが分かる。このような利用者に関しては、事前にトピックに対する興味が低いことが分かっているのであれば、主計画を短めに設計したり、そもそも興味が無いトピックを事前に回避するといった方策が有効だと考えられる。また、事前にトピックに対する興味が判定できなくても、パラ言語情報などを活用して、対話中に興味の有無を推定できれば、補足情報や付加情報を過剰に提供しないようなアプローチも考えられる。しかしながら、被験者 K のように、知らない情報が多いにもかかわらず、比較的良好に話す人もいることから、トピックに対する関心は高いが、欲しい情報が記事中に存在しなかったケースも考えられる。実際、7.1.8 で質問応答の誤り分析

を行っているが、記事中から回答できないような質問も多く存在していた。

以上の結果を整理すると、現状のシステムはトピックに興味があり、システムに好意的なユーザーに対しては有効に動作するが、トピックに興味がないユーザーに対しては過剰に情報を提供しがちであると言える。今後は、パラ言語情報を活用して興味を推定したり、主計画をユーザーごとに設計するなどの対策を検討する。

7.1.3 実験2:一問一答タスクの実験結果

実験1と同様に5点の文及び4点以上の文をどれだけ伝えられたかのグラフを図7.6、図7.7に、1点の文及び2点以下の文をどれだけ省けたかのグラフを図7.8、図7.9に示す。実験1の結果と比較すると、得点の傾向は類似しているものの、会話で補えた情報の量が実験2の方が少ないことが分かる。自由対話では、情報が提示されたら、その場で質問して回答を得たり、興味を示して補足情報を得たりすることができるが、一問一答では、要約提示時に気になった内容を記憶に留め、自分のターンがきた時点で、質問を具体化してから話す必要があるため、利用者の負担が大きい。例えば、自由対話では、「何それ」や「なんで」というだけで、文脈から回答が得られるが、一問一答では、「レアアースって何」や「なんで愛知でロボット産業が盛んなの」というように具体的に質問する必要がある。しかしながら、実験2においても発話内容は自由としているため、実験1のときのように感想を述べて補足情報を得たり、得られた回答に対して「何それ」などの簡易的な質問をして情報を得たりしているケースも見受けられた。

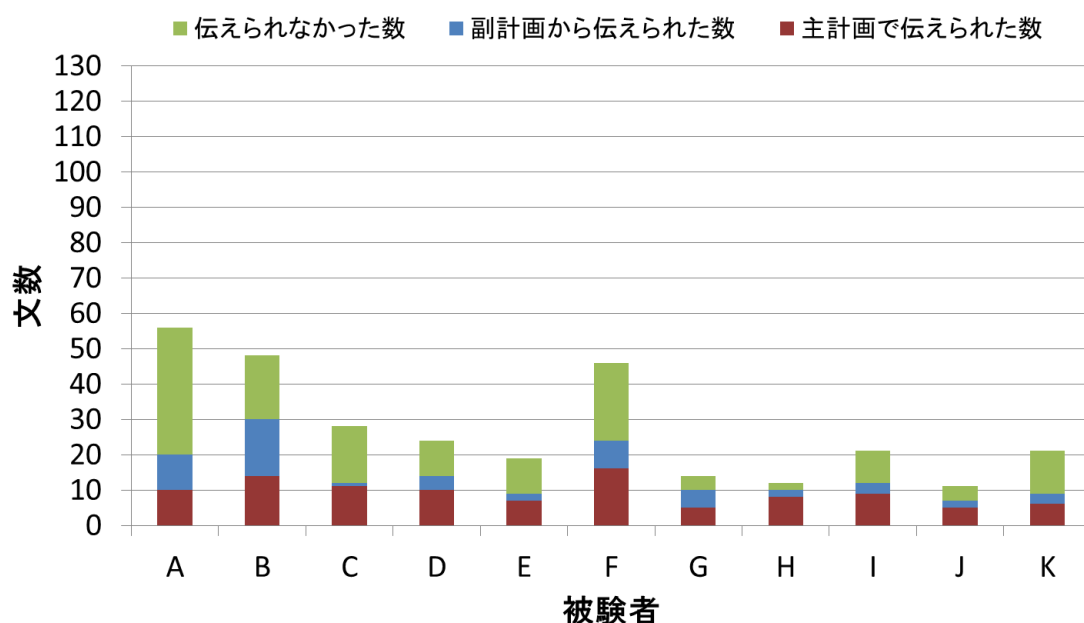


図 7.6: ユーザーが強く興味を持つ文（5点をつけた文）を伝えられたかどうかに関する分析結果（一問一答）

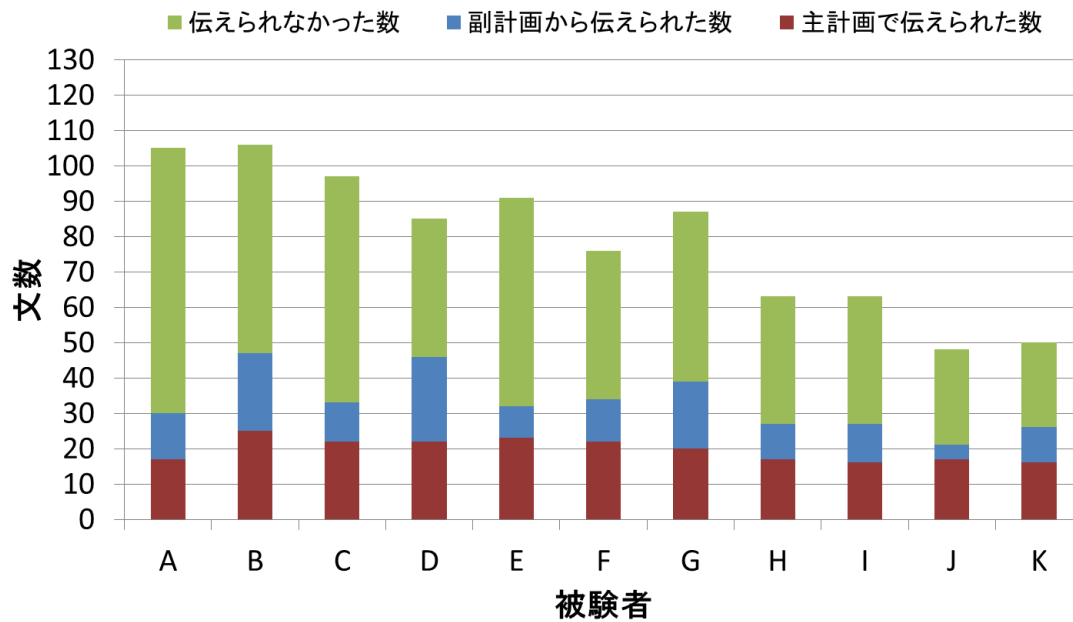


図 7.7: ユーザーが興味を持つ文（4 点以上をつけた文）を伝えられたかどうかに関する分析結果（一問一答）

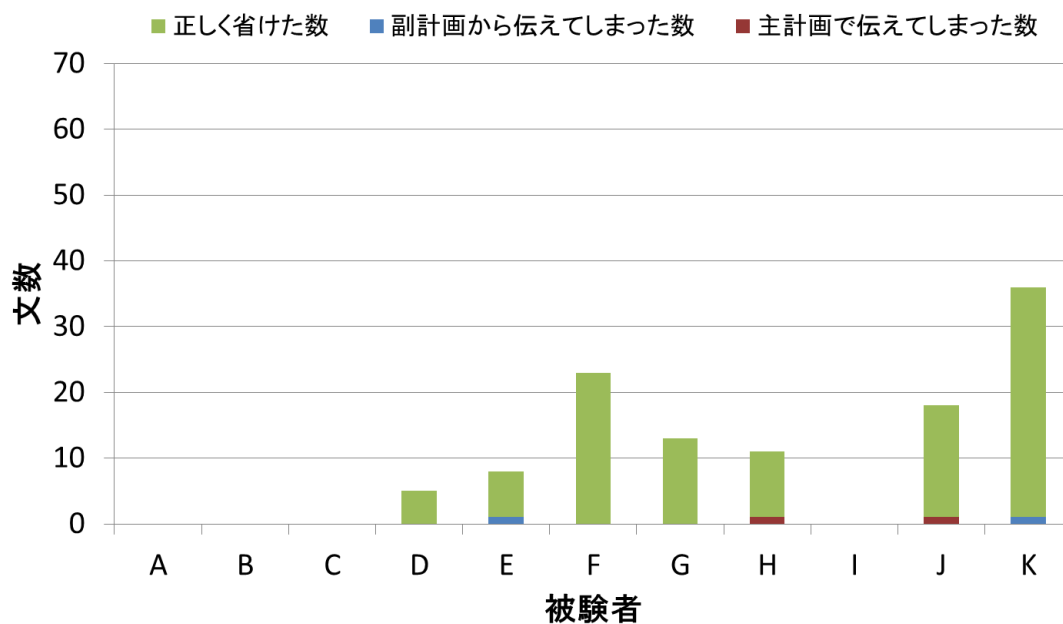


図 7.8: ユーザーが全く興味を持たない文（1 点をつけた文）を除外できたかどうかに関する分析結果（一問一答）

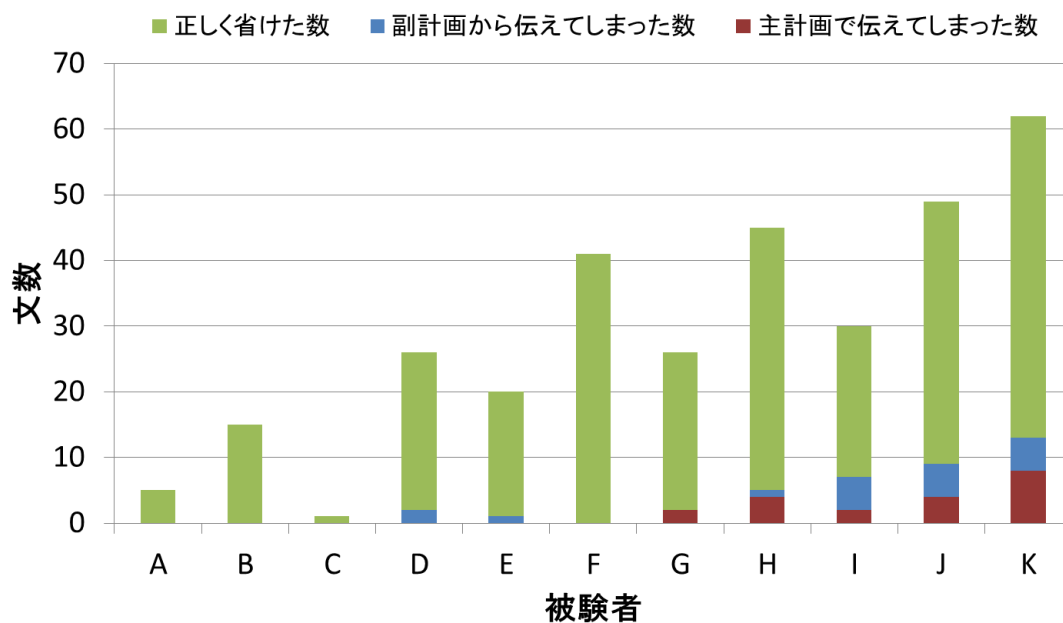


図 7.9: ユーザーが興味を持たない文（2点以下をつけた文）を除外できたかどうかに関する分析結果（一問一答）

ユーザーの発話の種類ごとの発話回数を図 7.10 に示す．実験 1 の結果（図 7.5）と比較すると，一問一答なだけに「質問・確認」の割合が高く，自由対話のときと発話数はほぼ同程度であるが，相槌や感想の数は激減していることが分かる．このことから，一問一答のシステムよりも本システムの方が多様な情報行動が出現しやすいことが分かる．

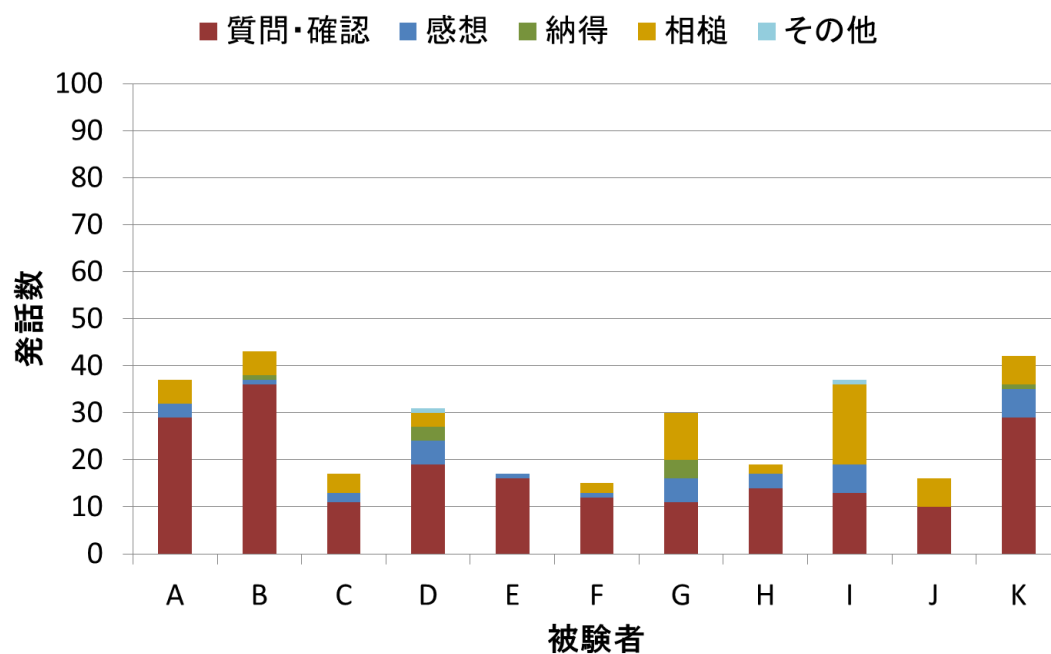


図 7.10: ユーザー発話の種類と回数（一問一答）

7.1.4 自由対話と一問一答の情報伝達効率の比較

自由対話タスクと一問一答タスクで、どれだけユーザーが欲しい情報を提示でき、どれだけ知らない情報を省くことができたかを総合的な観点で比較するために、情報伝達効率 (EoIT; Efficiency of Information Transfer) という尺度を提案する。EoIT は、提示できた4点以上の文の比率 R_c と除外できた2点以下の文の比率 R_r の調和平均である。

$$EoIT = \frac{2}{\frac{1}{R_c} + \frac{1}{R_r}} \quad (7.2)$$

全ての被験者の結果を総合して算出した EoIT の結果を図 7.11 に示す。図は、主計画で保証できた情報伝達効率と会話で実現できた情報伝達効率を示している。主計画で保証できた情報伝達効率に関しては、扱っているトピックの違いによって変化しうるが、自由対話と一問一答でその差がほとんどなかったことから、同様のトピックであったと言える。会話での情報伝達効率に関しては、自由対話の方が効率的な情報伝達を実現できていることが分かる。自由対話の方が一問一答よりも 1.12 倍高い情報伝達効率を示した。

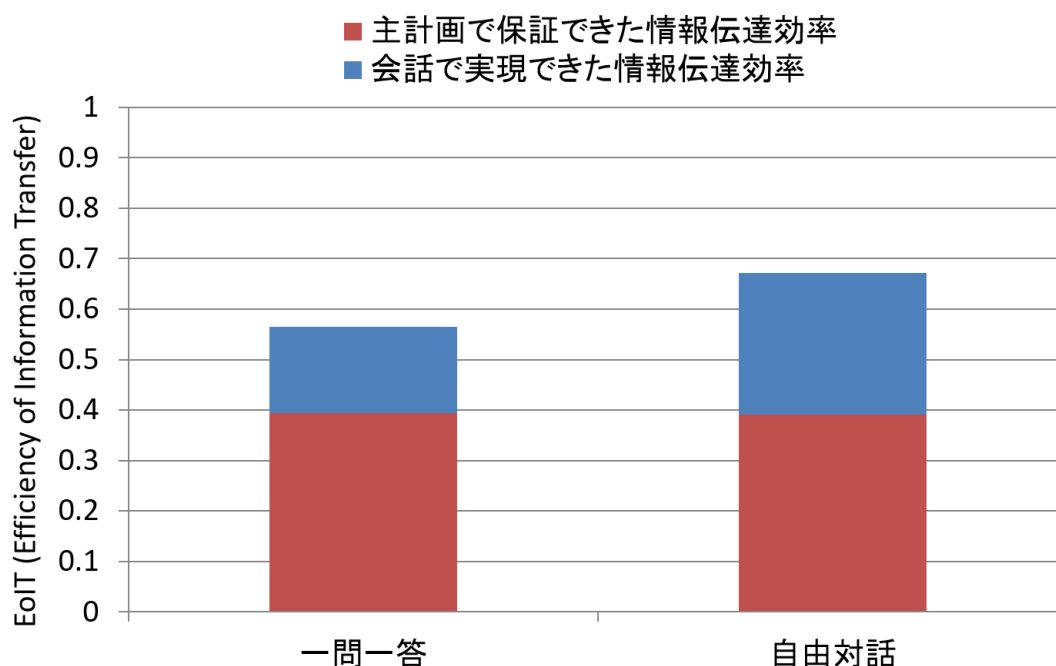


図 7.11: 自由対話と一問一答での情報伝達効率 (EoIT; Efficiency of Information Transfer)

7.1.5 主計画のサイズを変えたときの4点以上の文の被覆率と2点以下の文の除外率の変化

今回の実験では、主計画における重要文の数を5に設定した。しかしながら、主計画のサイズによって保証できる情報伝達効率が変わると考えられる。そこで、主計画のサイズ (K) を

変化させたときの、4点以上の文の被覆率と2点以下の文の除外率の変化を調べた。重要文の数を1から10まで変化させたときの、主計画の4点以上被覆率と2点以下除外率の変化、及び、自由対話と一問一答のそれぞれの会話タスクにおいて主計画と副計画で提示された情報に関する4点以上被覆率と2点以下除外率の値を図7.12に示す。ここで、主計画のみの対話とは、要約のみを聴かせることに相当する。主計画のサイズを変えるとユーザーの欲しい情報の被覆率と知らない情報の除外率はトレードオフの関係にあるが、会話によるインタラクティブな情報伝達により、顕著に被覆率と除外率の両方を改善できていることが分かる。質問応答と自由対話の比較では、質問応答の方がわずかに2点以下除外率が高いが、自由対話の方が4点以上被覆率は高く、欲しい情報をより多く引き出せていることが分かる。

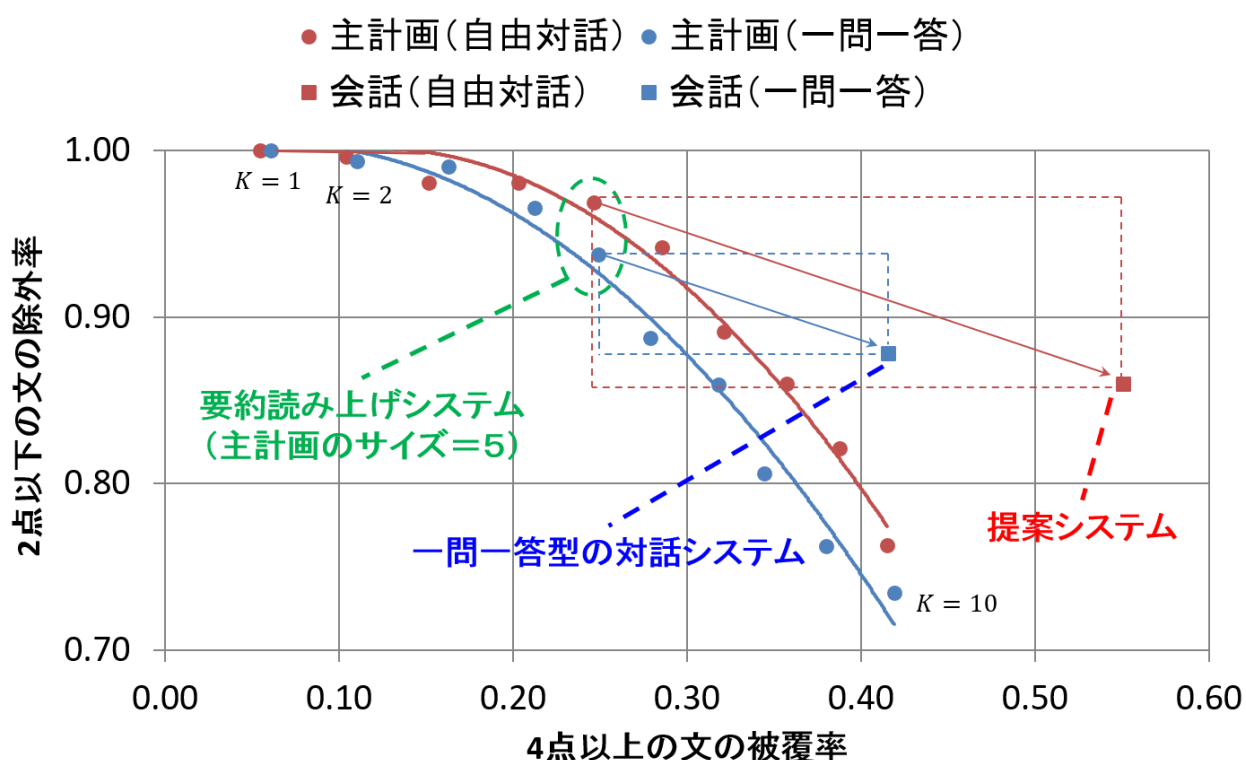


図 7.12: 自由対話と一問一答における主計画のサイズ (K) を変えたときの4点以上被覆率と2点以下除外率の変化

7.1.6 実際の会話例

実際の会話例を図7.13に示す。ここで、Sがシステム発話を表し、Uがユーザー発話を表す。ここで、主計画の内容は {S01, S02, S04, S05, S07, S09} である。特に、この会話で着目すべき箇所は、「何それ」(U03)の部分で、ユーザーは気になる単語がでてきた直後に「何それ」と質問して情報を得ていることが分かる。一問一答の使い方だと、主計画の内容を伝え終わるまで、

気になる単語を記憶しておき、自分のターンがきた時点で、具体的に質問しなければならない（例えば「プラグインハイブリッドって何」など）。しかしながら、発せられた音声はすぐに消えてしまうため（揮発性）、長時間記憶にとどめておくことは困難だと言える。また、仮に正しく反復できたとしても音声認識結果が正しいとは限らない。そのため、本システムのように「何それ」というだけで情報を瞬時に補完できる仕組みを持つことは極めて有用だと考える。

S01:	ヨーロッパで電気自動車への移行が加速してるそうだよ
U01:	へー
S02:	イギリス政府が、2040年までにガソリン車とかディーゼル車の販売を全面的に禁止するって発表したらしい
U02:	じゃあ、今の車はなくなっちゃうの
S03:	うーん、どうだろう
S04:	世界の石油消費の7割近くは自動車とかの輸送用が占めてるんだけど、環境対策として、脱石油が世界的な流れみたいだよ
S05:	日本政府も、2030年までに電気自動車とかプラグインハイブリッドとかの割合を
U03:	何それ
S06:	プラグインハイブリッドっていうのは、コンセントから差込プラグを用いて直接バッテリーに充電できるハイブリッド車のことだよ
U04:	ふーん
S07:	それで、そのプラグインハイブリッドとかの割合を5割から7割にする目標を掲げるんだって
U05:	できるの、そんなこと
S08:	それはちょっと分からない
S09:	でも、従来の燃料車向けの部品メーカーもたくさん残ってるから、大胆な政策変更はしにくいらしいよ
U06:	まあ、そうだよねー
U07:	いつまでにやるんだっけ
S10:	2030年だよ
U08:	2030年かー
S11:	自動車大国のドイツでも、2030年までにガソリン車とかの販売を禁止する決議が採択されたらしいよ
U09:	へー、そうなんだ
S12:	法制化までは至っていないんだけど、脱燃料車の機運が高まってるんだって
U10:	さすがヨーロッパ、進んでるなー
S13:	この流れはアジアにも及んでるそうだよ
U11:	そうなの
S14:	インド政府も2030年までに販売する車をすべて電気自動車にするっていう目標を掲げてて、中国でも類似の政策が打ち出されてるらしいよ

図 7.13: 実際の会話例（S がシステム発話，U がユーザー発話，{S01,S02,S04,S05,S07,S09} が主計画の内容）

7.1.7 Naive な副計画との比較

実験1及び実験2において副計画から伝達された文は、ユーザーの質問に対する回答として与えられたものやユーザーの興味・関心に応じて補足・付加されたものであるが、仮に同じ数の文を文の重要度に応じて与えた場合、または、ランダムに選択して与えた場合、どれだけユーザーの欲しい情報を与えられるかについて評価した。今、副計画から N 個の文が提示されたとする。このとき、主計画として選ばれた文を除き、残りの文からランダムに N 個選択した場合を「ランダム」、文の重要度でランキングし、上位 N 件を選択した場合を「重要度」とする。また、実際に会話で伝えられたものを「会話」とする。これら3つの手法に関して、選ばれた文の得点の総和を比較する。ここで、副計画をランダムに与えたものと文の重要度に応じて与えたものを Naive な副計画と呼ぶ。文の重要度は、文に含まれる単語（名詞、動詞、形容詞）の重要度の和とし、単語の重要度には、重要文抽出で使用した単語の重要度を用いる。

自由対話と一問一答のそれぞれに対する結果を図7.14, 図7.15に示す。「ランダム」と「重要度」の比較では、人によってはランダムの方が上回ることもあるが、平均的に「重要度」の方が高い得点を示している。「会話」との比較においては、全ての場合において「会話」が最も高い得点を示した。また、5点や4点の文が多い被験者に関しては、その差は僅差であるが、2点や1点の文が多い被験者に関しては、他に比べてわずかに差が大きくなっている。以上の結果から、ユーザーの発話内容を考慮しない Naive な副計画よりも、その時々でユーザーが示す興味に応じて提示する情報を決定する方法の方がユーザーが求めている情報をより多く提供できることが分かる。

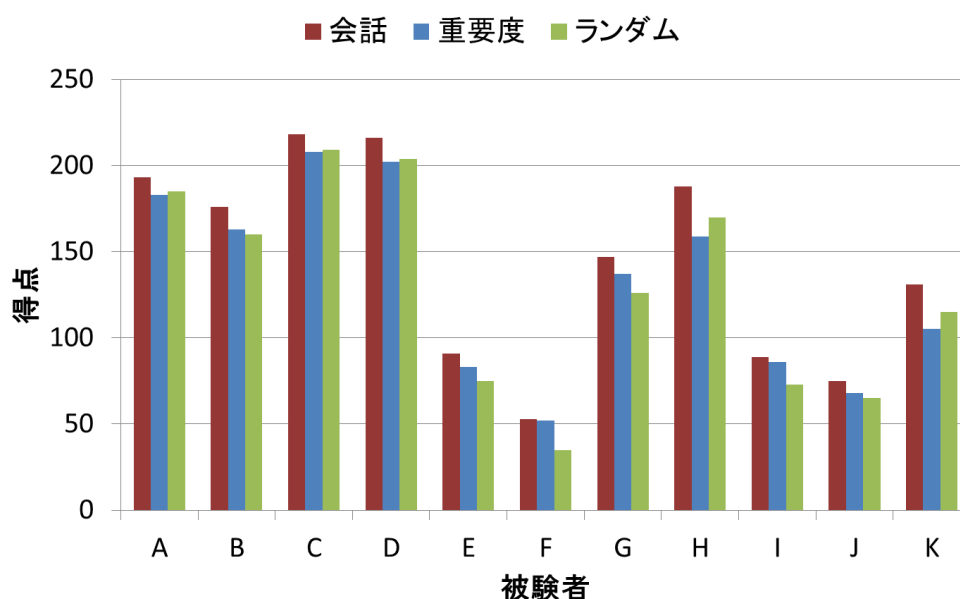


図 7.14: Naive な副計画との比較（自由対話）

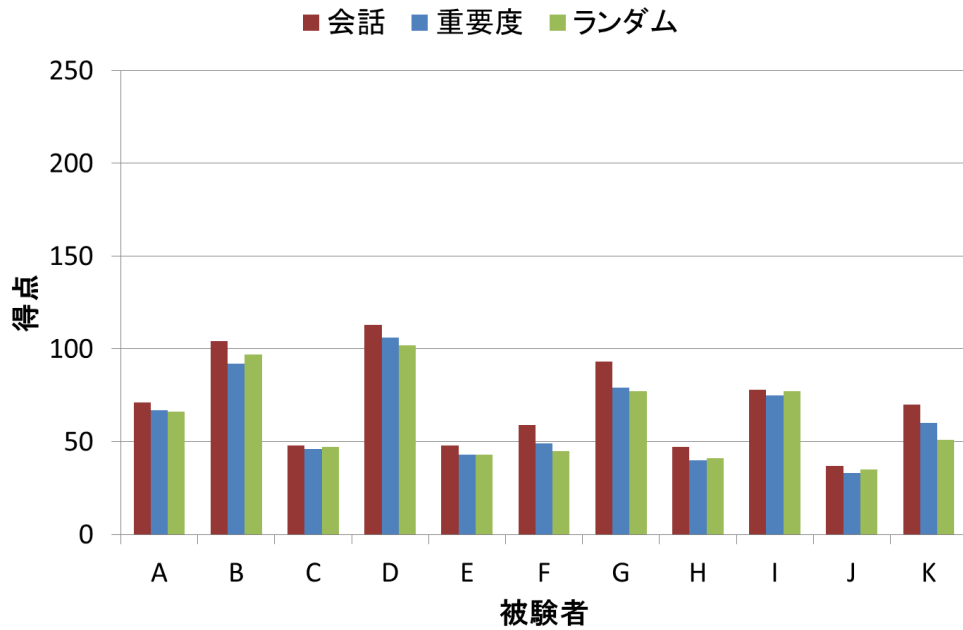


図 7.15: Naive な副計画との比較（一問一答）

7.1.8 質問応答の誤り分析

実験 1 と実験 2 におけるユーザーの質問に対して、システムがどれだけ適切に回答できたかについて評価を行った。質問応答では、記事内の情報だけでなく、辞書や Wikipedia の情報も使用して回答しているため、それらを区別して分析を行った。質問応答の種類を以下の 7 つに分類し、被験者ごとの割合を図 7.16 に、全体での割合を図 7.17 に示す。

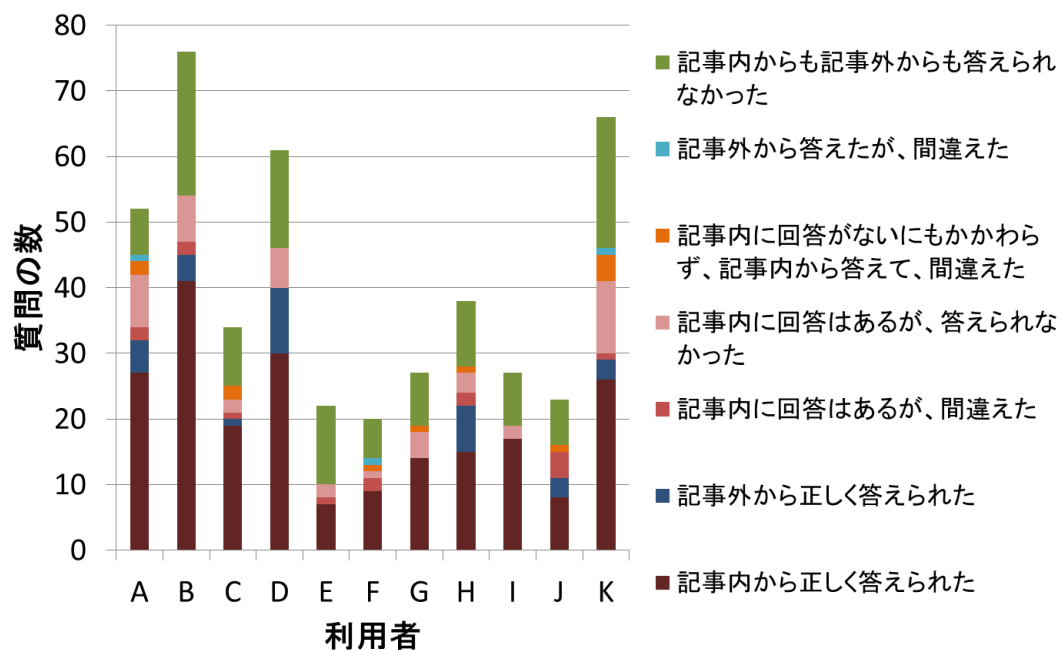


図 7.16: 被験者ごとの質問応答の種類と回数

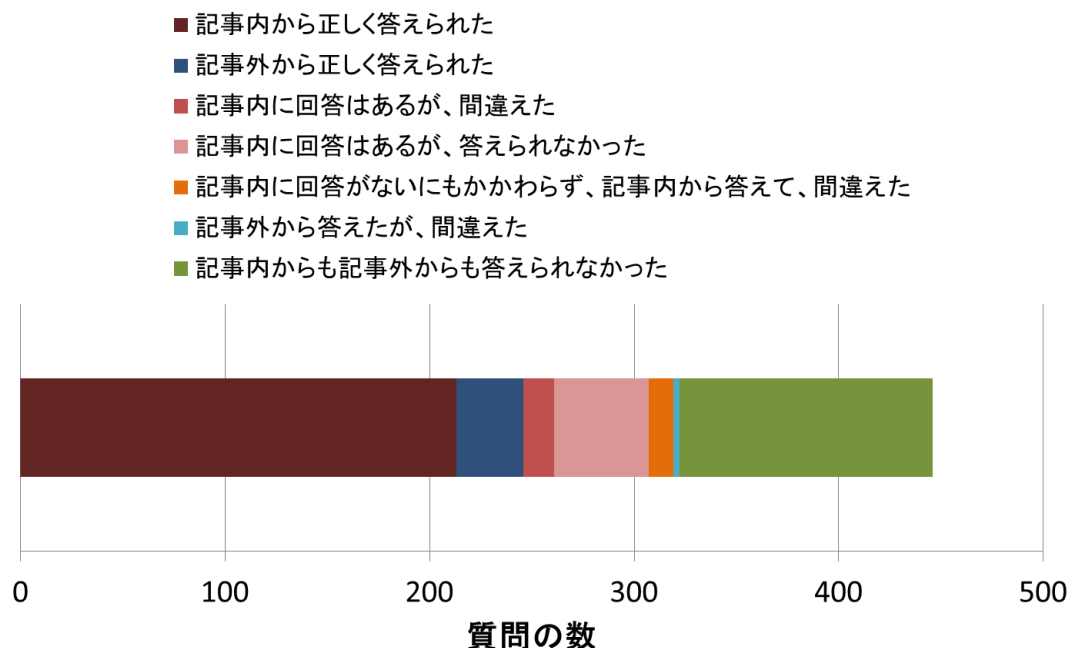


図 7.17: 被験者全体での質問応答の種類と回数

- 記事内から正しく答えられた：例えば、「それっていつ発売するの」という質問の回答が記事内にあり、それを正しく提示できた場合、これに属する。
- 記事外から正しく答えられた：例えば、「レアアースって何」という質問に対して、情報源として辞書などの外部知識を用いて答え、それが正しかった場合、これに属する。
- 記事内に回答はあるが、間違えた：例えば、あるサービスの利用料金についての質問（「それっていくらなの」）に対して、サービスの料金ではなく、工事費を答えてしまったなどといった、記事内に回答があるにもかかわらず、誤った回答を提示してしまった場合、これに属する。
- 記事内に回答はあるが、答えられなかった：例えば、「例えばどんな健康を管理するの」という質問に対して、記事内に回答があるにもかかわらず、「それはちょっと分からない」と答えてしまった場合、これに属する。
- 記事外から答えたが、間違えた：例えば、「日本の消費電力って全部でどれくらいなの」という質問に対して、Wikipedia の情報を利用して回答したが、製造業における電力消費率を回答してしまった場合などは、これに属する。
- 記事内に回答がないにもかかわらず、記事内から答えて、間違えた：例えば、「レアアースの消費量 1 位ってどこ」という質問に対して、記事内には 1 位の情報が書かれていないにもかかわらず、記事に書かれていた 2 位の「日本」を回答として提示してしまった場合などは、これに属する。

- 記事内からも記事外からも答えられなかった：記事内及び記事外の情報を用いても答えられなかったものは、これに属する。

このうち、「記事内から正しく答えられた」とものと「記事外から正しく答えられた」とものを正解として、被験者ごとに正解率を算出した結果を図 7.18 に示す。全体での平均正解率はおおよそ 55%であった。被験者の中では、被験者 D の正解率が最も高かった。この人の場合、「～って何」といった質問が多く、EDR 辞書や Wikipedia から獲得した用語の定義情報を用いた定義型質問応答でポイントを稼いでいた。今回の実験では存在しなかったが、予備実験において、多義性の問題により誤った回答を与えてしまうことがあった。例えば、「川崎フロンターレ」について聞きたいところを「川崎って」と質問して「川崎は、日本人の氏名だよ」という回答を得てしまうケースである。現状のシステムでは、記事中に「川崎フロンターレ」という単語が出現していれば、固有表現に関する文字列の包含関係に基づく共参照解析により「川崎＝川崎フロンターレ」と特定することで正しい回答を得ることができる。しかしながら、サッカーのニュース記事はそのことを前提として書かれているため、略称のみで正式名称が出現しないことがある。そのため、複数の意味が考えられるような単語が存在する場合、文脈も考慮し、その意味を正しく捉えられるような方策が必要である。

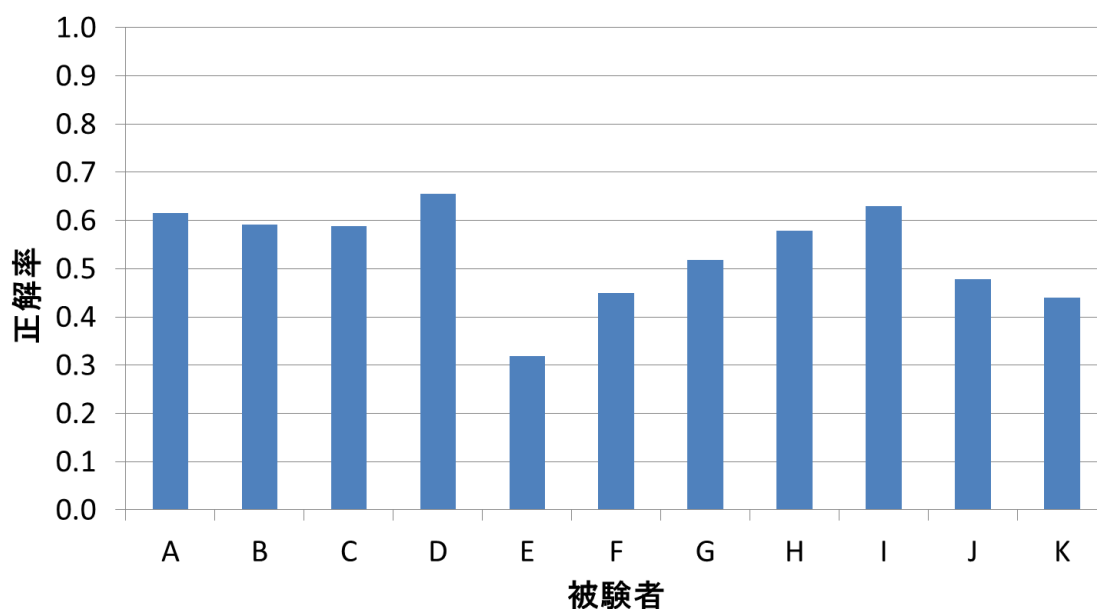


図 7.18: 被験者ごとの質問応答の正解率（全体の平均正解率は 55%）

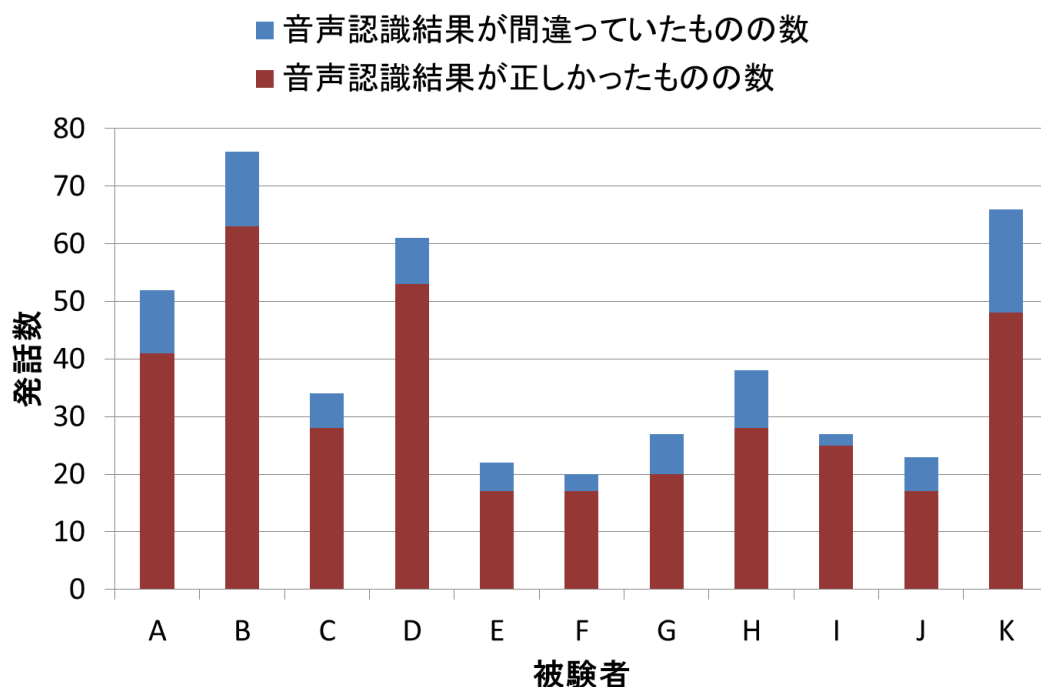


図 7.19: 質問応答に関する音声認識誤りの割合（全体の平均音声認識誤り率は20%）

正しく答えられなかったものの中で最も大きな割合を占めていた「記事内からも記事外からも答えられなかった」ものについてさらに詳しく分析する。原因の一つとして音声認識誤りが挙げられる。質問応答における被験者ごとの音声認識誤りの割合を図7.19に示す。平均でおよそ20%の音声認識誤りが見られた。しかしながら、この20%全てで質問応答が失敗していたわけではなく、正しく認識できた一部の情報を使って正しく回答できたケースも存在する。例えば、「レアアースって何に使うの」という質問の音声認識結果が「ペアーズって何に使うの」であったが、文脈的にレアアースの話題であったため、「何に使うの」という部分の情報だけで質問に正しく回答できていた。一方で、音声認識誤りが情報獲得で致命的になるケースも多く存在していた。例えば、「ネオジムって何」という質問の音声認識結果が「苗字ムって何」であったり、「どうして」が「投資で」、「何年」が「奈良根」であったがために回答できなかったものがよく見られた。今後の課題の一つとして音声認識誤りに対する頑健性の向上が挙げられる。

次に、音声認識結果が正しくても回答できなかった質問として、「日本は気候が不安定だけど、自然エネルギーは有用だと思いますか」や「太陽光発電は将来普及するの」といった意見や推測を求める質問、「私みたいな学生も受けれるの」といった利用者の属性情報を知らなくては回答できない質問、「レアアースとレアメタルってどう違うの」といった知識不足によって答えられなかった質問などがある。知識不足で答えられなかった質問に関しては、知識を拡充

することで対処できると考えられるが、意見や推測を求める質問に関しては、一つの答えが定まっているわけではない。一つの回答方針としては、一般的にはこう言われているというような情報を示すことだが、現状のシステムには実装されていない。

現状のシステムでは、分からない質問に対しては、「それはちょっと分からない」と答えるか、「なんで～なの」という質問に対して「なんでだろうね」と答えたり、先ほどの「太陽光発電は将来普及するの」のような質問であれば「どうだろうね」というように、回答をはぐらかすような返事をしている。しかしながら、分からないという返事ばかりになったり、毎回同じような言い回しで回答をはぐらかしてばかりでは、情報享受の快適さは損なわれてしまう。そのため、できるだけ多くの質問に対して正しい答えを提示できることが望まれる。

7.1.9 ユーザーの反応速度に関する比較

自由対話と一問一答でのシステム発話からユーザー発話までの平均時間を表 7.3 に示す。ここで、一問一答では、システム発話中に割り込んで発話することを許していないので、一問一答と同じ条件にするために、自由対話における計算では、バージインの発話を除いている。表 7.3 の結果から、自由対話の方が一問一答よりも即時的な反応が多いことが分かる。自由対話では、疑問に思ったことをその場で、しかも、「何それ」といった簡単な表現で質問できる。一方で、一問一答では、要約で伝えられた内容を記憶にとどめ、自分のターンがきたら、記憶を探りながら、具体的な質問を考えて発話する必要があるため、反応までにより多くの時間を費やしていると考えられる。しかしながら、主計画の発話の間隔を 2 秒としているため、ユーザーが焦って発話した可能性は否定できない。実際、ユーザー発話とシステム発話が衝突するケースが何度も観測された。現状のシステムでは、発話の衝突を検知したら、数秒間停止し、ユーザーの発話を待つようにしているが、発話の衝突が頻繁に起こると情報享受の快適さを損なうことになるため、今後の課題の一つとして動的な間の制御が挙げられる。

表 7.3: システム発話からユーザー発話までの平均時間（秒）

	A	B	C	D	E	F	G	H	I	J	K
自由対話	1.23	1.33	1.81	1.25	0.89	1.27	0.92	1.12	1.09	1.95	1.16
一問一答	4.01	3.18	3.24	3.28	3.93	3.01	3.35	3.42	5.01	3.37	4.45

7.2 継続利用可能性と眠気防止効果に関する評価

継続利用可能性と眠気防止効果に関する評価を行った。

7.2.1 実験設定

10人の被験者（20～30代 男性8名，女性2名）に対して本システムと読み上げシステムを利用させ，各システムについて「継続して使用したいか」「眠気防止に効果がありそうか」に関するアンケート調査を実施した．ここで，読み上げシステムとは，AITalkでニュース記事を最初から最後まで読み上げるシステムである．トピックには日経新聞のニュース記事を5つ使用した．継続利用可能性に関する評価は次の4段階で行った．

4pt 使いたい

3pt どちらかと言えば使いたい

2pt どちらかと言えば使いたくない

1pt 使いたくない

また，眠気防止効果に関する評価は次の4段階で行った．

4pt 効果がありそう

3pt どちらかと言えば効果がありそう

2pt どちらかと言えば効果がなさそう

1pt 効果がなさそう

7.2.2 実験結果

継続利用可能性に関する実験結果を図7.20に，眠気防止効果に関する実験結果を図7.21に示す．継続利用可能性については本システムが平均2.7点で読み上げシステムが平均1.9点であった．眠気防止効果については本システムが平均2.6点で読み上げシステムが平均1.6点であった．以上の結果から，机上環境において継続利用可能性と眠気防止効果に関する本システムの有効性が確認できた．

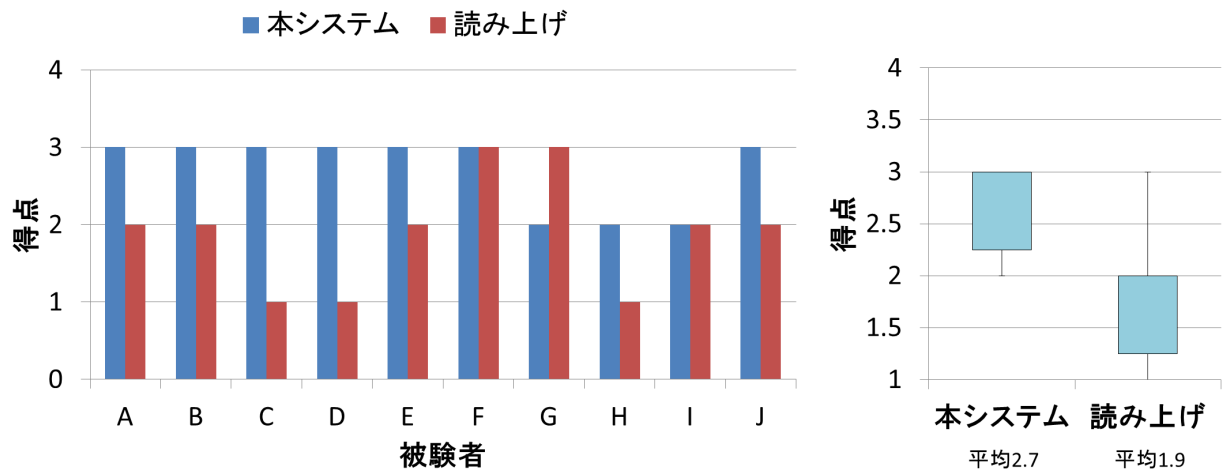


図 7.20: 継続利用可能性に関する実験結果

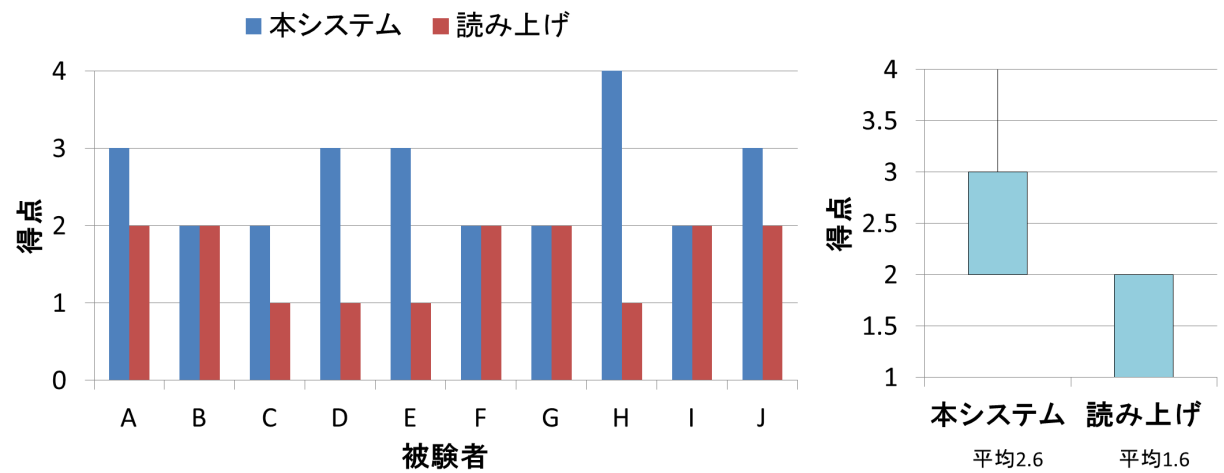


図 7.21: 眠気防止効果に関する実験結果

7.3 主計画のサイズを変えたときの情報伝達効率に関する評価実験

BERT に基づく要約モデルで生成した2種類の要約（要約 A と要約 B）を主計画として会話したときの情報伝達効率を評価する．DNN 音声合成システム（6 章参照）を使用し、「質問」に関する発話意図認識モデル（5 章参照）を組み込んだ会話システムで評価した．

7.3.1 実験設定

日経新聞のテクノロジー系のニュース記事8個（平均34.6文）について，4トピックずつ要約 A と要約 B を作成し，それぞれを主計画としたときの会話実験を8人の被験者に対して実施した．ここで，抽出する重要文 a の数を5，抽出する重要文 b の数を4に設定した（要約 A : 5 文，要約 B : 9 文）．また，本システムを制約なく使用する自由対話タスクと要約を述べた後に QA モードへ移行する一問一答タスク（例えば，吉野らのシステム [吉野 14, Yoshino 15a, Yoshino 15b]）について比較を行った．

被験者には，会話終了後にニュース記事を読ませ，各文について会話で伝えて欲しかったかどうかを5段階で評価させた（5点：伝えてほしい，4点：どちらかと言えば伝えてほしい，3点：どちらでもよい，2点：どちらかと言えばいらない，1点：いらない）．会話で提示できた4点以上の文の被覆率と会話で提示しなかった2点以下の文の除外率の調和平均から情報伝達効率を計算した．

7.3.2 実験結果

要約 A を主計画として会話したときの情報伝達効率の結果を図 7.23 に，要約 B を主計画として会話したときの情報伝達効率の結果を図 7.23 に示す．

この結果から，要約を伝えるだけのシステムや従来の一問一答型の対話システム（一問一答）と比べ，バージインを許容し即応性高く応答可能な本システム（自由対話）の方が高い情報伝達効率を実現できることが分かった．

また，今回の実験では被験者全体の傾向として知らない情報よりもほしい情報の方が多く見られたため，要約 A よりも要約 B の方が高い情報伝達効率を実現できた．一方で，要約を除き会話のみで実現できた情報伝達効率に着目すると要約 A に基づく会話も要約 B に基づく会話も情報伝達効率に大きな差は見られなかった．

主計画のサイズ（重要文 a の数）を変化させたときの4点以上被覆率と2点以下除外率の変化を図 7.25 に示す．要約のサイズを大きくすれば，ほしい情報の被覆率は向上するが，いらない

い情報の除外率が低下することが分かる．そのため，高い情報伝達効率を実現するためには，ユーザーに応じて適切な要約サイズを指定する必要がある．

合計抽出文数を10文に固定し，重要文タイプごとの抽出文数を変化させたときの情報伝達効率を図7.26に示す．この結果から，重要文aと重要文bの文数を適切に配分することで情報伝達効率の向上が期待できることが分かる．

今回の実験では，被験者に対して要約A，要約Bをランダムに割り当てて会話させたが，本来であれば，ユーザーのトピックに対する興味度に応じて要約Aと要約Bを使い分けるべきである．今後は，ユーザーとの対話履歴を活用し，ユーザーごとに適切な要約を作成する方法について検討する．

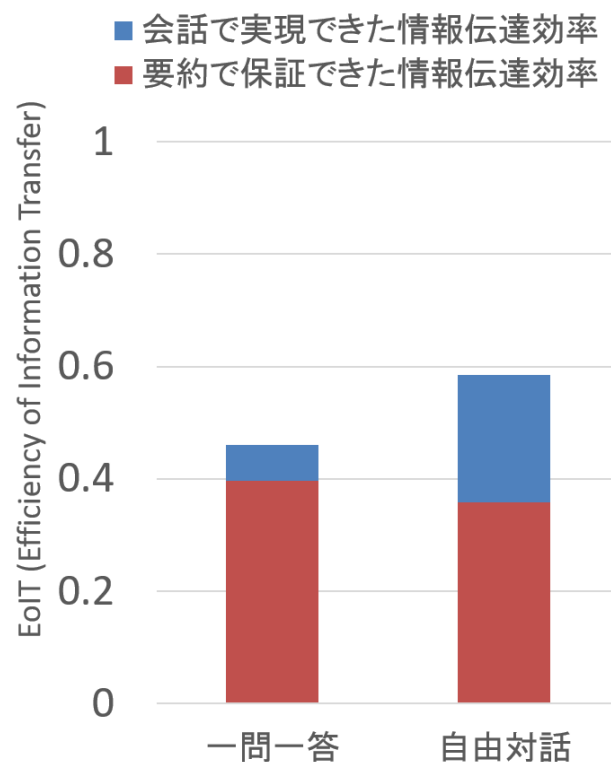


図 7.22: 情報伝達効率（全体）

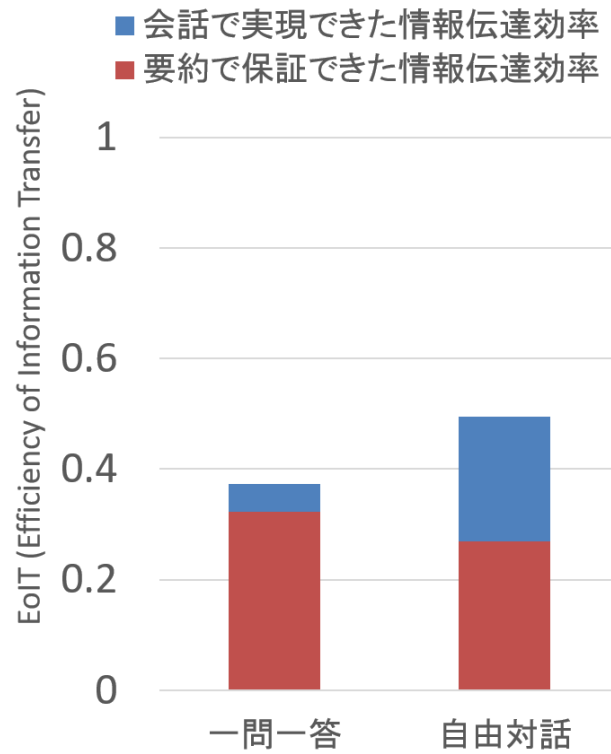


図 7.23: 情報伝達効率 (要約 A)

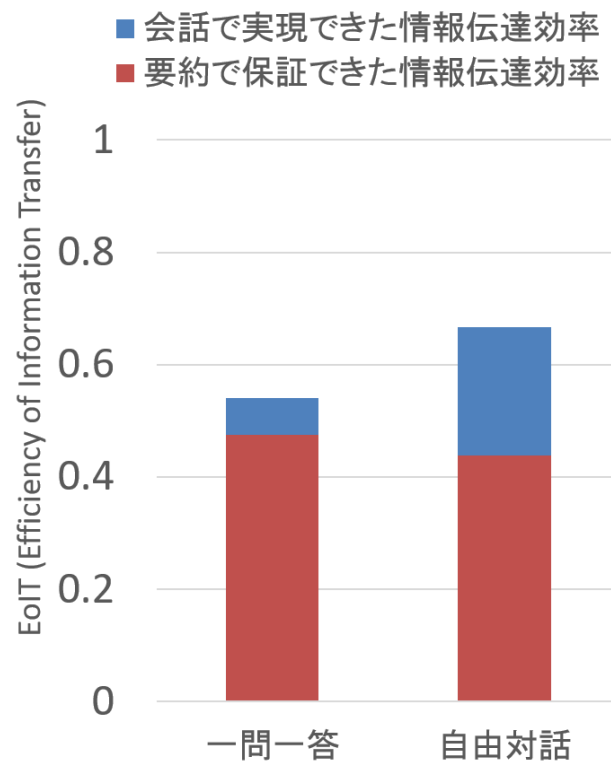


図 7.24: 情報伝達効率 (要約 B)

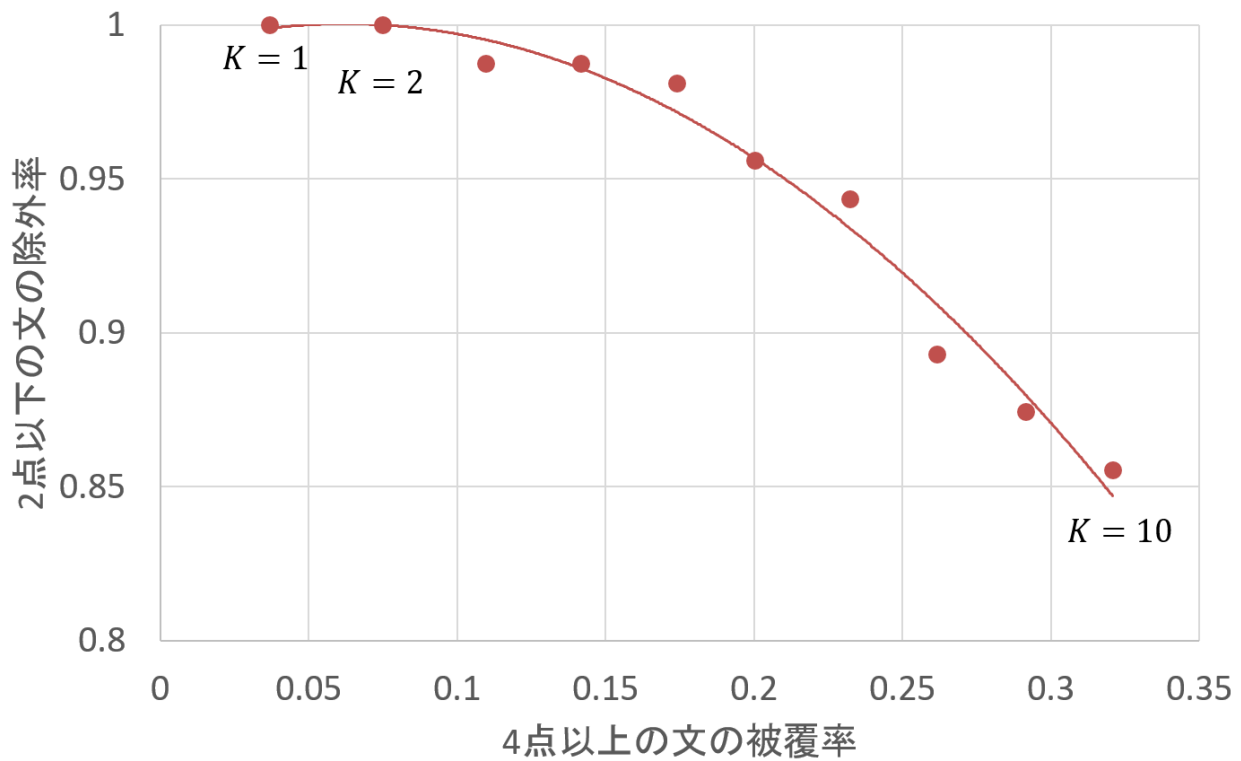


図 7.25: 主計画のサイズ（重要文 a の数 K ）を変化させたときの 4 点以上被覆率と 2 点以下除外率の変化

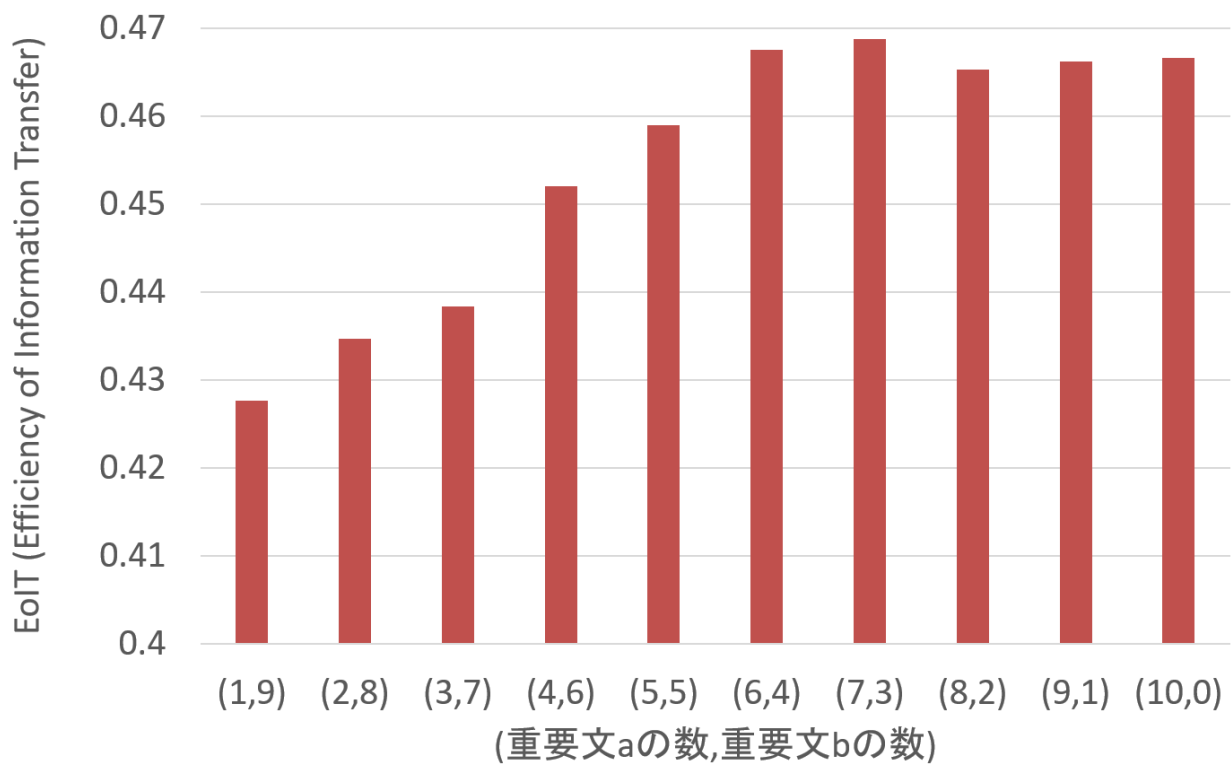


図 7.26: 重要文タイプごとの抽出文数を変化させたときの情報伝達効率（合計抽出文数は 10 で固定）

7.4 まとめと今後の課題

主計画と副計画からなるシナリオに基づいて、割り込みを許容しながら、質問に即座に回答できる点が本システムの特徴の一つである。このような即時応答機構を持つシナリオドリブンな提案システムと、要約を伝えた後にQAのモードに移行する従来の一問一答型の対話システムを比較評価した。その結果、情報伝達効率（EoIT; Efficiency of Information Transfer）という、提示できたほしい情報の被覆率と提示しなかったいらない情報の除外率の調和平均で定義される提案尺度において、提案システムの方が優れていることを確認した。また、本システムを継続して利用したいかどうかについて、アンケート調査を行ったところ、過半数が継続して利用したいと回答した。

情報伝達効率におけるアンケート結果から、ユーザーごとにほしい情報・いらない情報の傾向が多様であることが分かった。また、ユーザー発話の分析結果から、ユーザーごとにフィードバックとして返される発話の種類も量も様々であることが分かった。このことから、今後、情報伝達効率や継続利用可能性を改善していくうえで、ユーザーのタイプに応じてシステムの振る舞いを変えることが重要だと考えられる。そこで、今後は、ユーザーのタイプを早期に把握し、それに応じて有効な対話戦略が講じられるようなフレームワークについて検討する。

第8章 結論

快適な情報享受は、質問応答のような意図性の高い能動的な情報行動だけでも、文書閲覧・聞き流しのような意図性のない受動的な情報行動だけでも実現できず、これらの情報行動のモードを自由に行き来することにより達成される。本研究では、意図性の変化を捉えるのに音声対話が適しているという立場から、ニュース記事を例題としてまとまりのある情報を利用者が快適に享受するための枠組みについて検討し、意図性の異なる多様な情報行動を高頻度に即応性高く切り替えながら情報にアクセスできる会話システムを開発した。

提案システムは、主計画と副計画と呼ぶ、二種類の発話計画に沿って会話を進める。主計画は、ニュース記事など、伝達対象の文書を要約し、口語化することで作られるもので、文書の要点を伝えることを目的としている。ユーザーが受け身で聴いている限り、システムは主計画に従った情報伝達を行う。一方、副計画は、会話進行の各時点におけるユーザーの反応予測に基づいて、それらに対する応答を計画したもので、ユーザーの興味・理解度に応じた補足情報の提示を目的としている。ユーザーが情報要求を示してきた場合、システムは副計画に遷移し、補足説明を行う。これらの発話計画を事前に準備することで、目的とする即応性の高い円滑な会話インタラクションを実現した。システムの評価のために、情報伝達効率（EoIT; Efficiency of Information Transfer）という指標を提案した。EoITは、一連の会話によって、興味のある文を過不足なく効率的に伝達できたかどうかを図るための指標である。この指標を用いて評価を行った結果、要約のみの情報伝達や従来の一問一答型の対話システムに比べて、提案システムの方が情報伝達効率が顕著に高く、快適な情報享受が可能であることを確認した。

以下、要素技術に関するまとめと今後の課題について述べる。

8.1 主計画生成のための要約

第2章では、会話によるニュース記事伝達のための要約手法について説明した。本システムでは、円滑な情報伝達を実現するために事前に発話計画を作成しておく。発話計画は主計画と副計画から構成される。主計画はニュースの要点に関する発話計画であり、ユーザーが受け身の姿勢で聴いている限り、システムはこの主計画に沿って発話する。主計画はニュース記事を要約した内容を基盤に生成される。ここで、要約は重要文抽出、整列、文圧縮の3段階の処理

を経て行われる。重要文抽出は、文書から重要な文（核となる情報・見出し的な内容を含む文）を数文抽出する問題である。重要文抽出の手法として最大被覆モデルを利用した。最大被覆モデルは、制限要約長に収まる範囲で重要な単語をできるだけ多く被覆するような文集合を抽出する手法である。そのため、単語の重要度の与え方が肝となる。従来、単語の重要度としては、頻度や TF-IDF、ロジスティック回帰で推定した値などが用いられてきた。本研究では、BERT (Bidirectional Encoder Representations from Transformers) を基本構造とした双方向モデルにより、文脈を考慮して単語の重要度を推定する手法を提案した。要約コーパスを作成し、評価実験を行った結果、従来の文脈を用いない手法よりも、提案モデルで推定した単語重要度を用いたときの方が高い ROUGE-1 を示した。整列は、重要文抽出で抽出した文の提示順序を決める問題である。要約の対象が単一記事である場合は、記事の出現順で問題ないことを確認した。複数記事からの要約である場合は、ある記事の重要文と別記事の重要文前後の文章との類似度に基づいて前後関係を決める手法を提案し、有効性を確認した。文圧縮は、文自体を短く要約する問題である。文圧縮の問題を文節レベルの系列ラベリングの問題として定式化し、BERT に基づく文圧縮モデルを提案した。実験の結果、単純な双方向 LSTM モデルに比べて、提案モデルの方が高い精度を示した。

今後は、ユーザーとの対話履歴を活用し、ユーザーごとに適切な要約を作成する方法について検討する。

8.2 口語化

第3章では、会話によるニュース記事伝達のための口語化手法について説明した。ニュース記事は書き言葉で書かれているため、会話で伝えるのに適した話し言葉に変換する必要がある。本研究では、友達と会話するようなカジュアルな感覚で利用できるシステムを目指している。そこで、敬体（ですます調）ではなく、くだけた表現に書き換える。中性的なキャラクターを想定し、ルールに基づく書き換えを行った。口調は「情報のなわ張り理論」に基づき伝聞口調に書き換える。その他、同格の展開、連体修飾節の展開、長い重文の分割などの書き換えを行う。口語化によって生成された発話文の言葉遣いに違和感がないか確認するため、主観評価実験を行った。実験の結果、システムが生成した発話系列に対する評価結果が、人間が口語化したものに対する評価結果と同程度であることを確認した。

本研究では、友達のようなカジュアルな関係で情報を伝えることを意識して、口語化のルールを作成した。しかしながら、7章のシステム評価時に行ったアンケート調査では馴れ馴れしいというコメントもあった。そのため、今後は、丁寧な口調も導入し、ユーザーとの親密度に応じて、徐々にくだけた口調に変化させていくような制御方法について検討する。

8.3 副計画生成

第4章では、副計画の生成方法について説明した。要約で省かれた補足情報や予想される質問に対する回答を事前に副計画として用意しておくことで、ユーザーの情報要求に対して迅速な応答が可能となる。Wikipedia などを用いた用語の定義説明や、固有表現抽出結果を利用した日時、場所、人物などに関する回答計画、「ので」などの手がかり表現を利用した原因・理由に関する回答計画、含意・矛盾表現を用いた真偽判定、「何それ」といった抽象的な質問に対する回答計画を行う。質問応答として「何それ」に特化したシステムであることを被験者に伝え、応答の適切さについて評価したところ、68%程度の応答能力があることを確認した。副計画の事前設計による迅速な応答の有効性については7章の評価で確認した。

現状のシステムは、副計画における質問のカバー率と回答の精度が低いため、システムの信頼性という面で、ユーザーの継続利用可能性に関する評価に悪い影響を及ぼしていると考えられる。今後は、関連記事などを活用し、副計画を拡充させる方式について検討するとともに、質問応答の精度を向上させる。

8.4 ユーザー発話の意図理解

第5章では、システム発話の文脈を考慮した発話意図認識手法について説明した。より高い情報伝達効率を実現するためには、ユーザーからのフィードバックを正しく理解することが重要となる。従来、発話意図の識別は、F0などの韻律情報に基づいて行われてきた。しかしながら、フィードバックに込められる意図をユーザー発話の情報のみから判断するのは難しい。そこで、本研究では、システム発話の文脈を考慮した発話意図認識モデルを提案した。本システムを用いて収集した対話データに対して発話意図のアノテーションを行ったデータセットを使用し、評価実験を行った結果、システム発話の文脈を考慮したモデルの方が考慮しないモデルに比べて高い識別精度を示した。

今後は、ユーザーの違いを考慮できるように特徴量を工夫するなどして性能の改善を目指すとともに、表情などの視覚情報も考慮したマルチモーダルな発話意図認識手法についても検討する。

8.5 音声合成と間(ま)の制御

第6章では、談話構造上の発話の役割に応じたメリハリのある話し方を可能とする音声合成手法と間(ま)の推定手法について説明した。ニュースのようなまとまった量の情報を伝える会話システムでは、重要な情報がユーザーに伝わる話し方が求められる。この問題は、音声合

成の分野では、これまで一文内におけるプロミネンスの付与という形で扱われてきた。しかしながら、発話の強調に影響を与える韻律は、談話構造に応じても変化することが古くから指摘されており、一文より長い単位で韻律や間（ま）を制御する必要がある。そこで、本研究では談話構造上の文の役割を、核（最も伝えたい内容を含む文）、前置き（核を伝えるためのリード文）、補足（情報を補足する文）の3つに分類し、これらを DNN 音声合成の補助情報として利用することで、発話系列全体で韻律や間（ま）の制御が可能な音声合成システムを開発した。実験は次の3つの観点で行った。(1) 情報を伝える話し方としてのふさわしさに関する主観評価。(2) ニュースの理解度に関する客観評価。(3) 質問しやすさに関する評価。実験の結果、従来の1文単位の読み上げ音声で学習したモデルの合成音声よりも、情報伝達を意識させ段落単位で発話させた音声で学習したモデルの合成音声の方が、情報伝達のための発話としてふさわしいと感じるとともに、ニュースの理解度という面においても優れていることが分かった。さらに、発話の段落における位置や談話構造上の役割を補助情報に用いて合成することにより、メリハリのある聴きやすい話し方になることが分かった。また、発話の間（ま）に関しても、提案モデルで推定した間（ま）で発話した方が、短い単調な間（ま）で発話するよりも、聴きやすく頭に入りやすく、さらに質問しやすい話し方を実現できることが確認できた。

今後は、会話の活性度に応じて間（ま）の長さを変化させるなど、会話の状況やユーザーとの親密度に応じて、間（ま）や発話スタイルを制御する方法について検討する。

8.6 統合システム

第7章では、提案システムと従来システムの比較実験結果について説明した。主計画と副計画からなるシナリオに基づいて、割り込みを許容しながら、質問に即座に回答できる点が本システムの特徴の一つである。このような即時応答機構を持つシナリオドリブンな提案システムと、要約を伝えた後に QA のモードに移行する従来の一問一答型の対話システムを比較評価した。その結果、情報伝達効率 (EoIT; Efficiency of Information Transfer) という、提示できたほしい情報の被覆率と提示しなかったいらない情報の除外率の調和平均で定義される提案尺度において、提案システムの方が優れていることを確認した。また、本システムを継続して利用したいかどうかについて、アンケート調査を行ったところ、過半数が継続して利用したいと回答した。

今後は、システムの信頼性に関わる要約や質問応答などの質を改善していくとともに、ユーザーとの親密度という新しい軸を加えて、情報伝達効率や継続利用可能性を向上させる仕組みについて検討する。

付 録 A エラーハンドリング

A.1 序論・方法

音声認識誤りに起因する誤解の解消を目的としたエラーハンドリングの方法について検討する。

質問応答の正解率を低くしている原因の一つとして音声認識誤りが挙げられる。例えば、システムが「建設中のスウォッチ新本社は薄い木造シェルの限界に挑んだんだって」と発話した後、ユーザーが「スウォッチって何」と質問したとする。しかしながら、「そっちって何」と音声認識されてしまうことがある。人間であれば、「そっち」と聞こえたとしても文脈から「スウォッチ」のことであると判断できる。本研究では、このような文脈から推定できる音声認識誤りについて、単語のローマ字¹表記間での類似度（1 - 正規化編集距離）を計算し、類似度が閾値（e.g. 0.65）以上の単語に修正する（表 A.1）。一方、「木造シェルって何」という質問が「ソーシャルって何」と音声認識されてしまうような場合については、ローマ字レベルの類似度を持ってしても特定するのが困難であるため、このような文脈に存在しない単語についての質問がきた場合には、「え？ ソーシャル？」と聞き返すことでユーザーに音声認識誤りの可能性を示唆すると同時に訂正する機会を設ける。このようなエラーハンドリングの仕組みを導入することで質問応答の正解率が向上した（4.3 章参照）。

本章では、エラーハンドリングありのシステムとなしのシステムを被験者に使用させ、どちらが良かったかアンケート調査した結果について報告する。

表 A.1: ローマ字表記の単語間の類似度計算の例

音声認識結果	対象語	類似度
そっち (sotchi)	スウォッチ (suuotchi)	0.750
ボーリング (bo-ringu)	ボーイング (bo-ingu)	0.875
進行 (shinkou)	神鋼 (shinkou)	1.000

¹ローマ字化には pykakasi (<https://github.com/miurahr/pykakasi>) を使用した。

A.2 主観評価

3つのニューストピックについて被験者10名（男性8名，女性2名）にエラーハンドリングありのシステムとなしのシステムを使用させ，どちらの方が良かったかアンケート調査を実施した．ここで，質問させる用語は指定しており，まず，被験者はその用語についてわざと誤った発音で質問をする．その後正しい発音で質問する．このとき，エラーハンドリングありとなしでどちらが良かったか評価する．実際に評価で用いたシナリオを表A.2に示す．評価は次の4段階で行った．各被験者の評価結果を図A.1に示す．この結果から全ての被験者がエラーハンドリングありのシステムの方が良いと評価したことが分かる．実験後，被験者からは「エラーハンドリングがあった方がコミュニケーションをしている感じがして良い」「認識しづらかったという事実がフィードバックされた方がイライラが減ると思う」「誤った入力の結果をずっと聞かされるのは辛いので，エラーハンドリングがあった方が良い」というコメントが得られた．

4pt ありの方が良い

3pt どちらかというとなりの方が良い

2pt どちらかというとなしの方が良い

1pt なしの方が良い

表 A.2: エラーハンドリングの評価で使用したシナリオ

	システム発話と質問対象語（赤）	誤った質問	正しい質問
シナリオ1	神戸製鋼所のアルミ製部材の品質データ 改ざん を巡って・・・	解散って何	改ざんって何
シナリオ2	東京電力福島第1原発と同じ、 沸騰 水型原発に・・・	フットって何	沸騰って何
シナリオ3	1978年にできた 大規模地震対策特別措置法 は、 南海トラフ の・・・	何それ	南海トラフって何

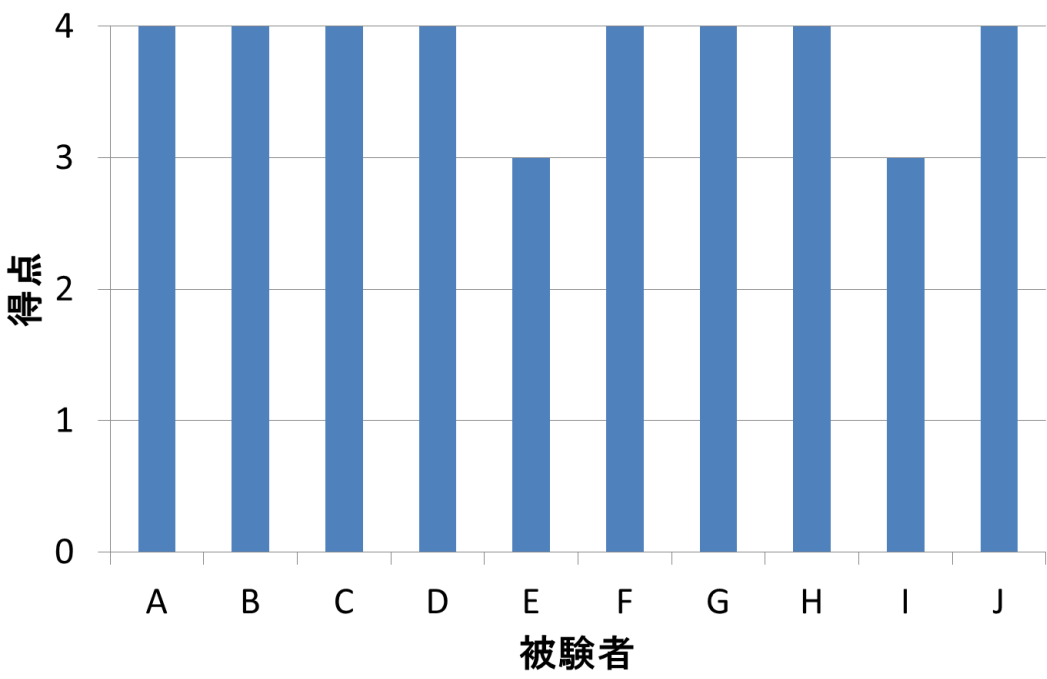


図 A.1: エラーハンドリング評価実験の結果

付 録 B トリビアの獲得

B.1 序論

トリビアを識別手法を提案する．また，ニュースを伝える会話システムにおいて，冒頭でトリビアを提示することの効果について検証する．

我々はニュース記事のようなまとまった量の情報を効率的に伝達する会話システムの開発を行っている．ここで「効率的」とは，伝達対象となる記事の中から，ユーザーにとって不要な情報を除き，必要な情報だけを伝えることを意味する．我々のシステムの特徴は，あらかじめ主計画，副計画と呼ぶ複数のシナリオを用意しておき，このシナリオに沿って会話を進めることで，リズムの良い会話を実現するうえで必須となる迅速な応答を可能としたところにある．主計画に沿って記事の要点となる情報を提示する傍らで随時ユーザーからのフィードバックを理解し，必要に応じて副計画に遷移して補足情報を提示する．このようにユーザーの興味や理解状態に応じて提示する情報を柔軟に切り替えながら会話を進めていく仕組みを持つ．一方で，高い情報伝達効率（EoIT; Efficiency of Information Transfer）を実現するにはユーザーからのフィードバックが必要不可欠である．

フィードバックに寄与する要素の一つとして，ユーザーのコンテンツに対する興味がある．従来より，コンテンツに対する興味を誘引する戦略として，会話主体のエンターテインメント（ゲーム，クイズショーなど）や教育などでトリビアが利用されてきた [Brown 84, Palincsar 84, ProMarketing Wizard 13, Voices Heard Media 13]．本システムでも，従来トリビアと呼ばれてきたものと同様，トリビアを「ある物事についての瑣末な知識であり，興味をそそるもの」と定義し，トリビアの提示がユーザーのコンテンツに対する興味を醸成する効果があると仮定する．コンテンツへの興味が高まることでフィードバックの質が改善し，情報伝達効率が向上することを期待する．

本研究では，トリビアか否かを識別するモデルを提案し，そのモデルを用いて Wikipedia からトリビアを自動獲得する．そして，ニュースに関連したトリビアを会話の冒頭で述べることにより，ユーザーのコンテンツに対する興味を誘引できるか，フィードバックの増加や情報伝達効率の向上に寄与するかという観点で評価する．

本章の構成は次の通りである．B.2 節で関連研究について述べる．B.3 節で提案するトリビアの識別手法について説明する．B.4 節でまとめと今後の課題について述べる．

B.2 関連研究

トリビアには、人の興味を引く性質があり、ゲームなどのエンターテインメントや教育現場などで利用されてきた [Brown 84, Palincsar 84, ProMarketing Wizard 13, Voices Heard Media 13]．例えば、Brown と Palincsar らは、読解力の指導においてトリビアは学生の注意をコンテンツに向けさせる効果があると指摘している [Brown 84, Palincsar 84]．

トリビアを自動で獲得する手法に関する研究やトリビアを対話システムの発話生成に応用した研究がある．新名らは、ユーザーを対話に引き込む効果を期待し、料理についての雑学を発話する対話システムを開発した [新名 17]．また、トリビアを含む文を「ある物事についての瑣末な知識であり、かつ人の興味を引くような内容を含む文」と定義し、主語の IDF や分散表現、主語と文中の名詞との類似度などを素性として文のトリビア度合いを回帰やランキング学習によって推定する手法を提案した [新名 18]．太田らは、対話を盛り上げるような発話の生成を目的として、Wikipedia の各文を TF-IDF、語の共起、文の長さによってスコア付けし、意外性のある文の抽出を行った [大田 09]．Prakash らは、固有表現の種類や Superlative Wordsなどを素性として Rank SVM を学習し、トリビアらしさのランキングを行った [Prakash 15]．Tsirel らは、トリビア度を人物とカテゴリの類似度の低さから計算される surprise とカテゴリに含まれる全ページ間の平均類似度の高さから計算される cohesiveness の積で定義し、人物の Wikipedia ページに付与されたカテゴリからトリビア度を計算する手法を提案した [Tsirel 17]．Fatma らは、(entity, predicate, object) の三つ組からなるデータを入力として、CNN でトリビアかどうかを識別する手法を提案した [Fatma 17]．また、人手で設計した特徴量と組み合わせることでより高い精度でトリビアを識別できることを示した．

本研究でも、従来トリビアと呼ばれてきたものと同様、トリビアを「ある物事についての瑣末な知識であり、興味をそそるもの」と定義する．また、Fatma らと同様にトリビアか否かを識別する 2 値分類問題として定式化する．本研究では、文の単語系列そのものを入力として、双方向 LSTM と self-attention [Lin 17] からなるモデルによって識別を行う．さらに、コンテンツに対する興味を誘引する効果が見込めるか、フィードバックの増加や情報伝達効率の向上に寄与するかどうかという観点で、ニュース伝達タスクにおいてトリビアを提示することの効果について評価する．

B.3 トリビアの識別手法

双方向 LSTM と self-attention [Lin 17] からなるモデルによってトリビアか否かを識別する手法を提案する．モデルの全体像を図 B.1 に示す．

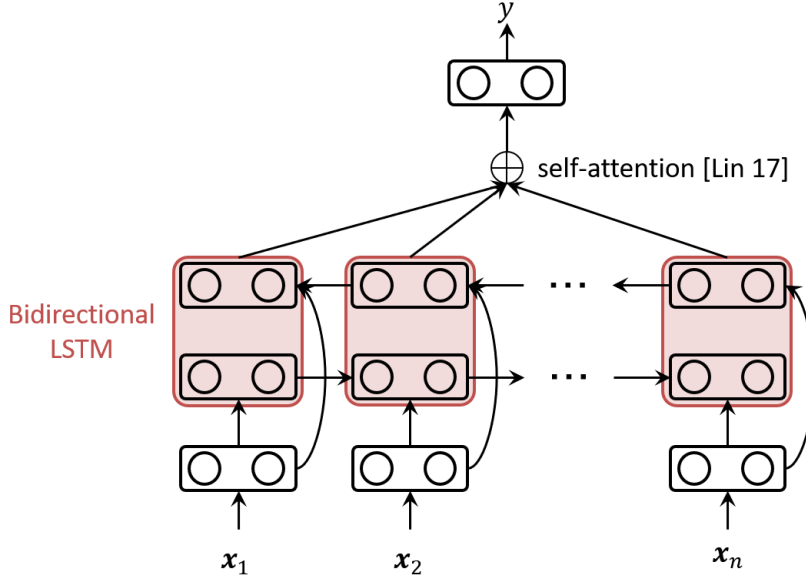


図 B.1: トリビア識別モデル

B.3.1 トリビア識別モデル

識別対象の文が n 個の単語系列 (x_1, x_2, \dots, x_n) から構成されているとする． x_i は i 番目の単語の one-hot ベクトルである．まず，各単語を埋め込み層で d 次元に圧縮し，得られた単語の埋め込み表現を双方向 LSTM に入力する．ここで，LSTM の隠れ層の次元を u とする．

$$e_i = W_{xe} \cdot x_i \quad (\text{B.1})$$

$$\vec{h}_i = \text{LSTM}(\vec{W}_{eh} \cdot e_i + \vec{W}_{hh} \cdot \vec{h}_{i-1}) \quad (\text{B.2})$$

$$\overleftarrow{h}_i = \text{LSTM}(\overleftarrow{W}_{eh} \cdot e_i + \overleftarrow{W}_{hh} \cdot \overleftarrow{h}_{i+1}) \quad (\text{B.3})$$

順方向 LSTM の出力と逆方向 LSTM の出力を連結し，それらを n 単語分並べて得られる行列を H とする．

$$h_i = [\vec{h}_i || \overleftarrow{h}_i] \quad (\text{B.4})$$

$$H = (h_1, h_2, \dots, h_n) \quad (\text{B.5})$$

この行列 H に対して attention を計算するが、長い文において複数箇所 (r 箇所) に注目できるように $r \times d_a$ 次元の行列 W_{s2} を掛ける (d_a はハイパーパラメータ).

$$A = \text{softmax}(W_{s2} \tanh(W_{s1} H^T)) \quad (\text{B.6})$$

得られた attention 行列 A を入力文の隠れ層の行列 H と掛けあわせて文の埋め込み行列 S を得る. この $r \times 2u$ 次元の行列 S を $2ru$ 次元のベクトル \mathbf{s} に変換し、文の埋め込みベクトルを得る.

$$S = AH \quad (\text{B.7})$$

$$\mathbf{s} = [S[1] || S[2] || \cdots || S[r]] \quad (\text{B.8})$$

最後にこのベクトルをシグモイド関数を活性化関数とする出力層に与え、トリビアである確率を計算する.

$$y = \sigma(W_{sy} \mathbf{s} + b) \quad (\text{B.9})$$

B.3.2 評価実験

Web から収集したトリビアを正例、Wikipedia からランダムに抽出してきた文を負例としてデータセットを作成し、トリビア識別モデルの性能を評価する.

B.3.2.1 実験設定

Web から収集したトリビアを正例、Wikipedia からランダムに抽出してきた文を負例 (非トリビア) とする. 負例を抽出するに当たり、トリビアらしい文が含まれないように負例は次の3つの条件を満たす文の中からランダムに抽出した. (1) 最初の段落でない, (2) 段落の先頭の文でない, (3) 記事タイトルを含まない. 作成したトリビアデータセットの統計を表 B.1 に示す. また、トリビアおよび非トリビアの例を表 B.2 に示す.

モデルのネットワーク構造として、埋め込み層の次元 d を 100, LSTM の隠れ層の次元 u を 100, attention の次元 d_a を 100, attention の数 r を 5 に設定した. 訓練セットにおいて頻度が 2 以下の単語は特殊記号 $\langle \text{unk} \rangle$ に置き換え、入力の語彙サイズを 4790 に制限した.

表 B.1: トリビアデータセットの統計

	訓練セット	開発セット	テストセット
正例	2500	100	126
負例	2500	100	126

表 B.2: トリビア／非トリビアの例

トリビア	非トリビア
童話「ウサギとカメ」で2匹が競争した距離は280m33cm未満	最初は馬の姿で現れるが、命じられれば人間の姿になる
アルゼンチンには、昔の丸の内線が走っている	また、自動車関連メーカーにはより安全な自動車の開発を促す
ミツユビナマケモノは首を最大270度回転させられる	ネオジム塩を成分とする鉱物の記事も含みます

B.3.2.2 実験結果

テストセットにおけるトリビアと非トリビアの識別精度を表 B.3 に示す．この結果から，提案モデルによって 97% と高い精度でトリビアか否かを識別できていることが分かった．

テストセットにおけるトリビアと非トリビアの埋め込み表現 \mathbf{s} の分布を多次元尺度構成法で次元圧縮し可視化した結果を図 B.2 に示す．この結果から，トリビアと非トリビアの境界がある程度明確に分かれるように学習できていることが分かった．

表 B.3: トリビア／非トリビアの識別精度（%）

Accuracy	「トリビア」 識別精度	「非トリビア」 識別精度
97.2	97.6	96.8

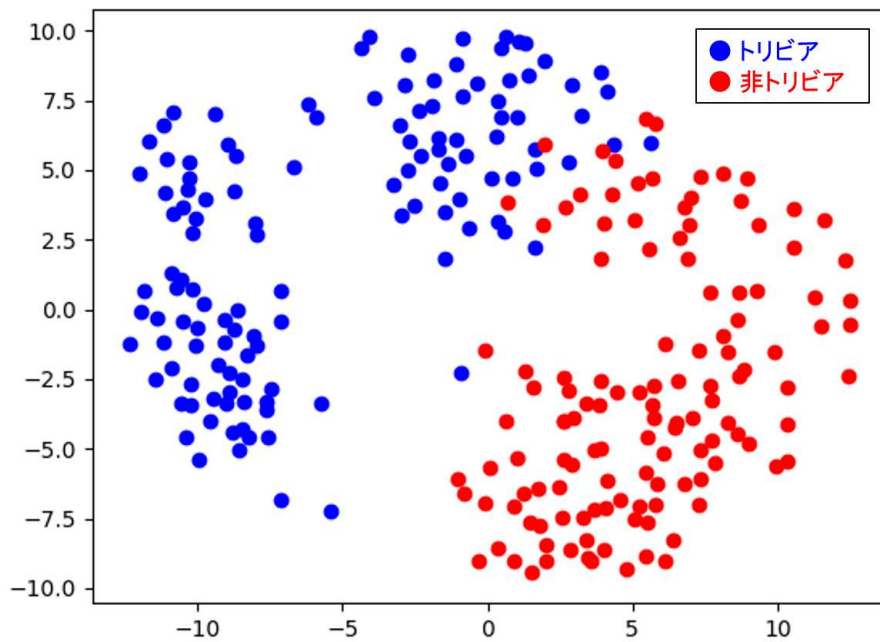


図 B.2: テストセットにおけるトリビアと非トリビアの埋め込み表現 s の分布を多次元尺度構成法で次元圧縮し可視化した結果

B.3.3 Wikipedia からのトリビアの獲得

学習した識別モデルを Wikipedia の記事に適用しトリビアを獲得する．ここでは，文脈依存性の高い文を回避するために以下の条件を満たす文に対して識別を行った．(1) 最初の段落である，(2) 段落の先頭の文である，(3) 記事タイトル+「は」を含む文である．Wikipedia の約 100 万記事に適用し，約 10 万文のトリビアを獲得した．

B.4 まとめと今後の課題

トリビアか否かを識別するモデルを提案し、97%と高い精度で識別できることを確認した。今後は、トリビアを提示する機能を会話システムに組み込み、以下の観点で評価する。

(1) コンテンツに対する興味を誘引する効果があるか

エンターテインメント [ProMarketing Wizard 13, Voices Heard Media 13] や教育 [Brown 84, Palincsar 84] の分野で言われているように、トリビアの利用が、コンテンツに対する興味を高めるのに効果があるか検証する。

(2) フィードバックの増加や情報伝達効率の向上に寄与するか

トリビアの提示が、フィードバックの増加や情報伝達効率の向上に効果があるか検証する。

(3) 会話の枕としての効果があるか:本題への流れを作る, ユーザーを聴くモードにさせる

落語の枕（マクラ）には、本題に入るための流れを作る効果があるとされている [広瀬 10]。トリビアの提示が会話の枕（マクラ）として効果があるか確認する。

(4) 親密度の向上に寄与するか

トリビアの提示がラポールの形成に寄与すると仮定し、短期的・長期的な期間においてトリビアの提示がラポールの形成・維持・崩壊のプロセス [Matsuyama 14, Zhao 14] に与える影響について分析を行う。

付 録 C オープンドメイン質問応答

C.1 序論

コンピュータが質問に対して適切に応答できることは言語を理解していることの証拠となる。そのため、質問応答は古くから言語理解研究において重要な位置づけを占めていた。

QA 研究への関心が高まったのは、情報検索技術に関する評価ワークショップ TREC (Text REtrieval Conference), 1999 年に質問応答を対象とする QA トラックが新たに設置されたことがきっかけの一つとなっている [Voorhees 01]。ここでの質問応答は、質問に対して分野や領域を限定しないことからオープンドメイン質問応答と呼ばれる。本研究でも扱うのも Wikipedia や Web コーパスを知識源としたオープンドメイン質問応答である。

質問応答の種類として、大きく分けてファクトイド型とノンファクトイド型がある [奥村 13]。ファクトイド型は単純な事実や出来事に関する質問であり、例えば、「世界で一番高い山は？」と言った質問である。2011 年にアメリカのクイズ番組 Jeopardy! でクイズ王を破り優勝した Watson が対象としていたのもファクトイド型である [金山 11a, 金山 11b]。一方で、近年研究が盛んになりつつあるのがノンファクトイド型で、これは、人物や物事の定義を尋ねる定義型、理由や原因を尋ねる why 型、形容や方法、手続きについて尋ねる how 型に分類される。また、近年、質問タイプを考えないタイプフリー型質問応答、または、Universal Question Answering (UQA) と呼ばれる質問応答研究も行われてる [水野 07a, Mizuno 07b, Soricut 06]。さらに最近では、SQuAD¹ [Rajpurkar 16, Rajpurkar 18] や TriviaQA² [Joshi 17], NewsQA³ [Trischler 17] といった大規模な QA データセットが公開されており、ニューラルネットを用いて回答を探索するアプローチが盛んに行われている [Devlin 18, Hu 18, Minjoon 17, Tan 16, Tran 18, Wang 18, Weissenborn 17, Yu 18]。日本語の QA データセット『解答可能性付き読解データセット』⁴も作成されている [鈴木 18]。

本章では、ファクトイド型質問応答と定義型質問応答、why 型質問応答について古典的なアプローチで取り組んだ結果についてまとめる。

¹<https://rajpurkar.github.io/SQuAD-explorer/>

²<http://nlp.cs.washington.edu/triviaqa/>

³<https://datasets.maluuba.com/NewsQA>

⁴<http://www.cl.ecei.tohoku.ac.jp/rcqa/>

質問応答システムは基本的に図 C.1 に示すようなモジュール構成をとる [東中 13a, 奥村 13, 佐々木 01]。以下, それぞれについて説明する。

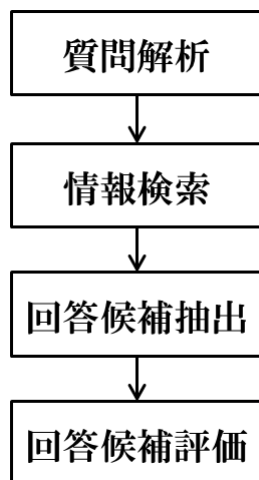


図 C.1: 質問応答システムの処理の流れ

C.2 質問解析

C.2.1 固有表現

一般に, ファクトイド型の質問応答では, 質問解析の段階でユーザーがどのような回答を期待しているのか大まかな目星をつける。その際に, 利用されるのが固有表現である。固有表現抽出は, もともとアメリカで開催された MUC (Message Understanding Conference) において, 情報抽出の重要な要素技術として着目され, 固有表現の種類として, PERSON, LOCATION, ORGANIZATION, DATE, TIME, MONEY, PERCENT が定められた。日本語の固有表現抽出は, 1999 年に開催された IREX (Information Retrieval and Extraction Exercise) という評価ワークショップのデータが標準となっており, MUC が定める固有表現に ARTIFACT を加えた 8 種類の固有表現を扱う。本研究でもこの 8 種類を対象としたファクトイド型質問応答を行う。

なお, 近年では 8 種類による分類では粗すぎて回答候補を絞り込めないことが指摘されており, 関根らは拡張固有表現という 200 種類の分類方法を提案している [Sekine 02, 関根 07, Sekine 08]。拡張固有表現タグ付きコーパスを作成する研究も行われており [橋本 04, Higashinaka 12], 『現代日本語書き言葉均衡コーパス』(BCCWJ)⁵のコアデータおよび『CD-毎日新聞' 95 データ集』⁶の新聞記事に対して拡張固有表現を付与したコーパスが言語資源協会 (GSK) で配付さ

⁵https://pj.ninjal.ac.jp/corpus_center/bccwj/

⁶<http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html>

れている⁷。また、現在、Wikipedia 構造化プロジェクト「森羅」の一環として Wikipedia の項目を拡張固有表現に分類する試みが行われている [関根 18a, 関根 18b]。

固有表現認識の手法は様々なものが提案されている [Lafferty 01, 笹野 08, 風間 08a, Kazama 08b, 岩倉 11, Lample 16]。入れ子にタグ付けされた固有表現を認識する手法も提案されている [Sohrab 18]。最近の固有表現認識に関するデータセットや手法については Yadav らのサーベイ論文にまとまっている [Yadav 18]。

固有表現をオープンドメイン質問応答に活用した研究の一例として、貞光らは、少量の固有表現集合からブートストラップ法で大規模な固有表現集合を獲得し [貞光 12]、獲得した固有表現間の関係を係り受け関係をもとに求め [Hirano 10]、オープンドメイン質問応答に活用した [貞光 13]。

本研究では KNP⁸の解析結果として得られる IREX の 8 種類の固有表現を用いる。

C.2.2 質問解析器の構築

古くは、質問解析は有限状態オートマトンなどを利用したルールベースの手法が多かった（例えば、FASTUS [Appelt 93] や Alembic [Aberdeen 95, Vilain 96]）。近年では SVM などの機械学習を利用した手法が多く用いられる [Hasan 16, 佐々木 04]。本研究では、識別器に決定木を利用する集団学習の手法である RandomForest を分類器として用いる。分類するクラスは IREX の固有表現 8 種類に why 型と定義型を加えた 10 種類である（表 C.1）。

表 C.1: 本研究で扱う質問タイプ

PERSON	人名	日本の初代総理大臣は誰ですか？
LOCATION	地名	日本の首都はどこ？
ORGANIZATION	組織名	大隈重信が設立した大学は？
ARTIFACT	固有物名	湯川秀樹は何を受賞しましたか？
DATE	日付	アメリカの独立記念日はいつ？
TIME	時間	競技かるたの暗記時間は何分ですか？
MONEY	金額	2013年のサマージャンボ宝くじの2等の賞金は？
PERCENT	割合	太陽の質量のうち何%がヘリウムですか？
WHY	原因／理由	どうして人工知能は人間を超えられないのか？
DEFINITION	定義	一般相対性理論とは？

⁷<https://www.gsk.or.jp/catalog/gsk2014-a/>

⁸<http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

C.2.2.1 素性

形態素解析にはJUMAN⁹を使用し、構文解析にはKNPを使用する。素性としては、JUMANの形態素情報（単語、品詞大分類、品詞細分類、カテゴリ、ドメイン）を利用する。以下、これら形態素情報を単語属性と呼ぶ。鈴木らは質問タイプの同定において単語属性のN-gramが有効であることを確認した [鈴木 03]。ここでは、さらに単語属性を係り受け経路上に制限して、単語属性の連続2-gram、非連続2-gramを取ったものを素性に加える。

C.2.2.2 評価実験

質問タイプごとに質問事例を作成し、RandomForestで学習し、10分割交差検定で評価を行った。RandomForestの設定条件を表C.2に示す。実験結果を表C.3に示す。エラー分析の結果、「ORGANIZATION」と「LOCATION」、「DATE」と「TIME」の間の誤りが多かった。

表 C.2: RandomForest の設定条件

決定木の数	100
ランダムに選択する説明変数の数	$\sqrt{\text{説明変数の総数}}$
乱雑さの評価基準	ジニ不純度

表 C.3: 質問解析の性能

質問タイプ	精度	再現率	F 値	データ数
ARTIFACT	91.2	92.9	91.9	322
DATE	93.1	93.4	93.2	558
LOCATION	92.9	92.6	92.7	598
MONEY	95.6	92.9	94.2	254
ORGANIZATION	94.8	85.2	89.5	455
PERCENT	96.6	95.3	95.9	449
PERSON	90.0	97.5	93.6	648
TIME	96.0	95.0	95.5	399
WHY	91.1	90.5	90.8	230
DEFINITION	96.8	94.2	95.5	220

⁹<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

C.3 ファクトイド型質問応答

ファクトイド型は「世界で一番高い山は？」のような単純な事実や出来事を尋ねる質問である。ファクトイド型の代表的なシステムとして、堀らの SPIQA [Hori 03], Harabagiu らの VAQA [Harabagiu 02], 佐々木らの SAIQA [佐々木 01] と SAIQA-II [佐々木 04], Moldovan らの PowerAnswer [Moldovan 07a], IBM の Watson [金山 11a, 金山 11b], NTT のしゃべってコンシェル [東中 13b]¹⁰ などがある。NTT は、しゃべってコンシェルの拡張機能として、知識 Q&A サービスも提供している [内田 13]。

SPIQA は、大規模音声認識エンジンを用いて質問に答えるシステムである [Hori 03]。SPIQA には、曖昧性解消モジュール DDQ (Deriving Disambiguating Queries) が備わっており、質問文を係り受け解析した結果、修飾語がないものや検索結果に頻出する表現がある場合、質問文を曖昧であると判定する。

VAQA (Voice-Activated Question Answering) は、英語の音声質問応答システムである [Harabagiu 02]。VAQA では、音声認識結果としてもっともらしい候補を複数挙げ、典型的な質問文パターンと照らし合わせて、不適切な候補を除外するという方法をとっている。

SAIQA (System for Advanced Interactive Question Answering) は、ルールによる質問解析、TF-IDF による段落検索、FST による固有表現抽出 [佐々木 99] などを組み合わせたシステムである [佐々木 01]。さらに、質問解析精度や固有表現抽出精度を改善するために、固有表現抽出、質問解析 [鈴木 03], 回答評価 [Suzuki 02] に SVM を用いたシステムが SAIQA-II である [佐々木 04]。

PowerAnswer は論理に基づく推論によって回答を生成するシステムである [Moldovan 07a]。PowerAnswer では、COGEX [Moldovan 03, Moldovan 07b] という定理証明器が用いられており、TREC-2007 のファクトイド問題で高い性能を示した。なお、COGEX は含意関係認識 (RTE; Recognizing Textual Entailment) でも良い成績を収めている [Tatu 07]。

C.3.1 回答候補の抽出

質問応答システムの多くは、多様な質問に回答できるように全文検索技術を利用している。全文検索エンジンとしては、Namazu¹¹ や Lucene¹², Indri¹³ などがある。しかしながら、より QA に特化した検索エンジンを開発する IR4QA (Information Retrieval for Question Answering) という研究分野もある。

¹⁰https://www.nttdocomo.co.jp/service/shabette_concier/

¹¹<http://www.namazu.org/index.html.ja>

¹²<http://lucene.apache.org/>

¹³<https://www.lemurproject.org/indri/>

データベースとして MySQL¹⁴を使用し、検索エンジンとして全文検索エンジンである Mroonga¹⁵を使った。データベース上には文そのものと、検索用文字列、KNP で固有表現抽出を行いタグ付けを行った固有表現タグ付け文、および形態素情報を付与した形態素タグ付け文をデータベース上にアップロードした。また、質問タイプごとにデータベースを用意することで検索の際の探索範囲を抑えている。検索用文字列は、文の形態素情報だけでなく、文の属する記事のタイトルや同義語・異表記情報（『基本的意味関係の事例ベース』¹⁶）、上位語情報（『上位下位関係抽出ツール』¹⁷で Wikipedia から抽出したもの）、類（C.5.4 節参照）、含意動詞（『動詞含意関係データベース』¹⁸）を含めることで検索ヒット率を高めている。

C.3.2 回答候補のランキング

現在の実装はコサイン類似度を利用した簡単な方法である。まず、回答候補文と質問文のコサイン類似度をとりランキングし、さらに上位 10 件で再ランキングする。この際、同じ回答が含まれていたら、スコアを加算する。さらに、質問に含まれる固有名詞が回答候補文に含まれている場合、その名詞からの距離（単語数）に応じて重み付けしている。簡易的な方法であるため、評価時間は短縮できるが、Bag-of-Words による類似度評価であるため、構文的・意味的関連性が考慮されていない。そのため、このことに起因する回答誤りは多く、十分な性能とは言いがたい。質問応答の例を表 C.4 に示す。

実際に会話で活用する際は、基本的に「～だよ」の口調で回答するが、確信度に応じて自信のなさを言葉で表現する（e.g. 「たぶん～」 「おそらく～」）。今後は、拡張固有表現に対応させ、より詳細な質問に回答できるようにする。

¹⁴<https://www.mysql.com/jp/>

¹⁵<http://mroonga.org/ja/docs/characteristic.html>

¹⁶<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-9>

¹⁷<https://alaginrc.nict.go.jp/hyponymy/>

¹⁸<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-2>

表 C.4: 固有表現を用いた質問応答の例

質問	回答候補	スコア
オードリー・ヘプバーン主演の映画は？	<u>おしゃれ泥棒 2</u>	0.831
	<u>ティファニーで朝食を</u>	0.495
	<u>シャレード</u>	0.414
	<u>おしゃれ泥棒</u>	0.413
	<u>マイヤーリング</u>	0.294
マレーシアの首都は？	<u>クアラルンプール</u>	3.04
	ジョホール王国	0.449
	日本	0.442
	郊外セランゴール州セパン	0.407
東日本大震災が起きたのはいつ？	<u>2011 年 3 月 11 日</u>	2.98
	<u>2011 年度</u>	2.00
	<u>2011 年</u>	1.69
東日本大震災は何時何分に起きたの？	<u>14 時 46 分 18 秒</u>	0.890
	<u>14 時 46 分</u>	0.529
	14 時 48 分 31 秒	0.379
	12 時 20 分	0.275
	午後 3 時 25 分	0.255
日本で最も大きな湖は？	霞ヶ浦	0.428
	<u>琵琶湖</u>	0.428
	北海道	0.426
	汽水湖	0.426
	ミクリガ池	0.225

C.3.3 意味的關係を用いた質問応答

NICTの音声質問応答システム「一休」[Saeger 12]を参考に、意味的關係を用いた質問応答手法について検討した。意味的關係とは、二つの名詞間に成り立つ關係のことであり、意味的關係の獲得にはSaegerらの手法を利用した[Saeger 09a, Saeger 10]。これは、クラス制約付きパターンを用いた關係インスタンスの獲得方法であり、構文パターンにクラス制約を設けることでジェネリックパターンの多義性を解消している。

(例) : $X : c_i$ の $Y : c_j$

→ 癌 [病氣] の治療薬 [薬] ... 予防關係

→ 京都 [地名] の金閣寺 [名所] ... 所在地關係

意味クラスの獲得には、風間らが導入した確率的クラスタリング法を利用する[風間 08a, Kazama 08b]。これにより、15万名詞を1000クラスにクラスタリングし、意味的關係として因果關係、予防關係、材料關係の3種類に関して獲得を行った。

表 C.5: 獲得した意味的關係の例

因果關係	予防關係	材料關係
ストレス ⇒ 躁うつ病	保湿力 ⇒ ひび割れ	カルシウム ⇒ 骨
過労 ⇒ 脳腫瘍	シートベルト ⇒ 転倒	カフェイン ⇒ コーヒー
食べ過ぎ ⇒ 肥満症	放射線治療 ⇒ 悪性リンパ腫	血小板 ⇒ 血液
不眠 ⇒ 交通事故	香水 ⇒ ワキガ	半導体 ⇒ CPU
タバコ ⇒ 肺がん	鎮痛剤 ⇒ 痛み	錐体細胞 ⇒ 眼球

意味的關係を用いた質問応答の手続きを以下に示す。

1. 質問対象語の特定

疑問詞 (e.g. 何, 誰) に係る名詞または末尾が疑問形の名詞節である場合の名詞を質問対象語とする。

(例) : 癌に効く薬は? → 質問対象語=“薬”

癌に——

効く——

薬は?

2. 述語項構造解析

KNP を用いて述語項構造解析を行い，2 項からなる構文パターンを作成する

(例) : 効く に:癌 が:薬 $\rightarrow \langle A$ が B に効く \rangle A =薬, B =癌

ただし，質問対象語 (A) を項に含める．

3. 言い換えパターンの生成

ALAGIN の日本語パターン言い換えデータベースを使用して，構文パターンに対する言い換えパターンを抽出する．

(例) : $\langle A$ が B に効く \rangle

$\langle B$ に効果がある A \rangle

$\langle B$ に効く A \rangle

$\langle B$ に良いとされる A \rangle

$\langle B$ に効果的な A \rangle

$\langle A$ が B に効果がある \rangle

$\langle A$ が B に良い \rangle

⋮

4. 意味的關係獲得時の構文パターンとマッチングを行う

5. マッチした構文パターンが持つ名詞対の一方と B でマッチングを行い， A を獲得する

(例) : 癌に効果的な抗癌剤

6. 回答候補のランキングを行う

現在は，意味的關係獲得時のスコアに IDF で重み付けした値を利用している．IDF は出現する文書が少ない単語ほど大きな値になる．これにより，「ウイルス \Rightarrow 風邪」よりも「インフルエンザウイルス \Rightarrow 風邪」を重視したランキングが行え，抽象的な回答を避けることができる．

意味的關係を用いて回答できる質問のパターンはそれほど多くはないものの，固有表現で対処できない質問に対する回答方法として有用である．対話システムで使用する場合は，文脈情報を加えて回答候補をランキングすることで，より適応的な回答ができると考えられる．

表 C.6: 意味的關係を用いた質問応答の例

質問	回答候補	スコア
癌に効く薬は？	マイトマイシン	0.0348
	抗ガン剤	0.0325
	治療薬	0.0317
醤油には何が含まれていますか？	香気成分	0.0118
	メラノイジン	0.00946
	アミン類	0.00768
ストレスを抑制する物質は何ですか？	オキシトシン	0.00383
	生理活性物質	0.000198

C.4 why 型質問応答

人間にとって、理由が知りたいという欲求は根本的なものであり、理由を知ることが物事の本質を理解するうえで重要である。why 型質問応答は、「どうして～？」、「なぜ～？」のような原因や理由を尋ねる質問に答えるものである。質問応答の中でも特に近年需要が高まっているのが why 型である。代表的なシステムとして原田らの Metis [Harada 07] や NTT の NAZEQA [磯崎 08, 磯崎 09] などがある。また、NICT の WISDOM¹⁹ [木俵 12, Mizuno 16, 水野 17, Tanaka 13] にも why 型（および how 型）の質問応答機能がある [Hashimoto 14]。

諸岡らは、森本らの why 型質問の回答分析 [森本 04] に基づいて人手で作成したパターンにより回答を探すシステムを作成した [諸岡 06, 諸岡 07]。渋谷らは、事実文と理由文の位置関係を考慮した回答抽出を行う why 型質問応答システム RE:Why を開発した [渋谷 07]。原田らは、意味解析システム SAGE [原田 01, 原田 02, 前澤 04, 梅澤 08, 川口 09] を用いて意味グラフベースで質問文と知識文を照合することで回答を抽出するシステム Metis を開発した [Harada 07, 久保田 08, 西岡 09, 高山 10, 原田 11]。東中と磯崎らは、EDR コーパス²⁰の意味解析結果を利用し、機械学習によって理由表現のパターンを自動獲得し、そのパターンを用いて回答するシステム NAZEQA を開発した [Higashinaka 08a, Higashinaka 08b, 磯崎 08, 磯崎 09]。

why 型質問応答の目的は、質問の内容を結果として持つ事柄の原因を応答することである。そのため、why 型質問応答の研究は、因果関係獲得に関する研究と密接に関連している。

¹⁹<https://wisdom-nict.jp/#top>

²⁰http://www2.nict.go.jp/ipp/EDR/JPN/TG/Doc/EDR_J09a.pdf

Selfridge らは、エンジンに関する説明文を入力されると、ある部品がどうなると他の部品がどうなるかを示すシステム CMACS (Causal Model Acquisition System) を開発した [Selfridge 85]. Joskowicz らは、戦艦の装置に関するエラー報告書を入力とし、装置の状態間の因果関係を理解するシステムを開発した [Joskowicz 89]. Kontos らは、物事の説明文から因果関係を抽出するシステムを開発した [Kontos 91]. Garcia は、フランス語の文章を対象として、causal verb が見つかった場合、その文に含まれる名詞句を抽出し、名詞句中に動作やプロセスを表す動詞派生名詞があれば、そのうち二つを因果関係にあると判定するシステム COATIS を開発した [Garcia 97]. Khoo らは、半自動で作成した構文パターンによって、原因と結果を表す名詞句ペアを獲得する手法を提案した [Khoo 00]. Chang らは、名詞句間および文間の因果関係を判定するのに、教師なし学習を適用し、80%のF値を得た [Chang 04]. 乾らは、「ため」を手がかり表現として因果関係知識を獲得し、4つのタイプ (cause, effect, precondition, means) に分類した [乾 04]. Chan らは、証券取引に関するニュース記事から、どの銘柄のどのような動きがどのような結果を引き起こしたかについて、三つ組のテンプレート (原因プロセス, causal verb, 結果プロセス) を抽出するシステム SEKE (Semantic Expectation-based Knowledge Extraction methodology) を開発した [Chan 05]. Do らは、述語と述語、述語と名詞、名詞と名詞といった単語ペアの関連度を自己相互情報量などに基づいて計算し、イベント間の因果関係を判定する手法を提案した [Do 11]. Oh らは、正規表現で原因・結果表現を含む文を取り出し、CRFで原因部と結果部のチャンキングする手法を提案し、複数文の文間関係を考慮した理由表現の獲得を行った [Oh 13]. Pichotta らは、Encoder-Decoder モデルを用いてスクリプト知識の獲得を行った [Pichotta 16]. Granroth-Wilding らは、述語項構造ペアを用いて、ニューラルネットワークによって因果関係の強さを推定する手法を提案した [Granroth-Wilding 16]. Sharp らは、因果関係の推定に役立つ分散表現を学習する手法を提案した [Sharp 16]. その他、因果関係の自動獲得手法については文献 [Asghar 16] にまとめられている.

本研究では、Oh らの手法 [Oh 13] を参考に原因と結果を一文に含んでいるものを対象として why 型質問応答システムを構築した.

C.4.1 原因部と結果部のチャンキング

C.4.1.1 因果関係候補文の抽出

正規表現を用いて、原因／理由表現にマーキングを行う。正規表現のルールを表 C.7 に示す。
ただし、〈定型句〉 = { 理由, 原因, 要因, 引き金, おかげ, せい, わけ }
以降、この正規表現を有する文を因果関係候補文と呼ぶ。

表 C.7: 原因／理由表現獲得のための正規表現ルール

(指示詞 の) ? ため 助詞	ため, のために, そのため
ので	ので
こと (から で)	ことから, ことで
(から ため) 判定詞	ためだ, からだ
指示詞? 〈定型句〉 (助詞 判定詞) +	理由は, 原因だ, この理由から

C.4.1.2 形態素素性

素性として、形態素情報（基本形、品詞大分類、品詞細分類、カテゴリ、ドメイン）を利用する。これらに対して、対象形態素の前後 5 個を考慮する。

(例) : 雨が降ったので、家に帰った。

[[降る], [降る, ので], [が, 降る], [が, 降る, ので], ...]

[[動詞], [動詞, 助動詞], [助詞, 動詞], [助詞, 動詞, 助動詞], ...]

C.4.1.3 構文素性

因果関係候補文のうち、マーカーを含む節に対する係り受け関係から、各形態素に構文ラベルを付与する。

表 C.8: 構文ラベル

c-marker	原因／理由マーカーを含む節
child	c-marker に係る節
parent	c-marker が係る節
subtree-of-child	child に係る部分木
subtree-of-parent	parent に係る部分木
others	それ以外

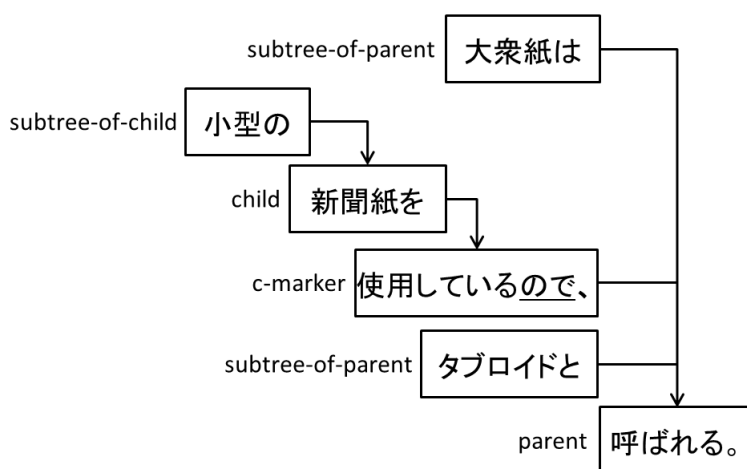


図 C.2: 構文情報のタグ付け例

C.4.1.4 CRF によるチャンキング

まず，人手で原因部と結果部のタグ付けを行う．

(例) : <c> 雨が降った </c> ので， <e> 家に帰った </e>.

そして，IOB ラベルを表 C.9 のように設定し，CRF で学習し，因果関係候補文に対して CRF によるチャンキングを行う．

表 C.9: 因果関係獲得のための IOB ラベル

B-C	原因部の開始
I-C	原因部の内側
B-E	結果部の開始
I-E	結果部の内側
O	外側

C.4.2 回答候補の抽出

ファクトイド型と同様で、データベースとして MySQL を使用し、検索エンジンとして全文検索エンジンである mroonga を使う。アップロードするデータは、文と形態素タグ付け文、原因部と結果部の文、および、それらの形態素タグ付け文である。ただし、原因部と結果部の述語は終止形に直し、また、原因部の末尾は「ため」「のため」となるように修正した。

C.4.3 回答候補のランキング

質問文と回答候補文のコサイン類似度と質問文と回答候補の結果部のコサイン類似度の和に基づいてランキングする。why 型質問応答の例を C.10 に示す。今後は機械学習によるランキング方法を検討する予定である。

回答候補のランキング手法として磯崎らも Oh らも SVM を使用しているが、Oh らは素性として、評価極性が有効であることを確認した [Oh 12]。これは、ネガティブな結果の原因もネガティブであるというような仮定に基づいたものである。また、Oh らは橋本らが提案した活性／不活性極性 [橋本 12a, Hashimoto 12b, Hashimoto 12c, 佐野 13] を用いることで、質問文と候補文の含意関係認識を近似している。

なお、活性／不活性極性は、「助詞＋用言」(e.g. を使う, を防ぐ) をテンプレートとして、これらの接続関係（例えば、「ので」は順接関係、「けど」は逆接関係）とシードテンプレートの極性をもとに物理学のスピンモデルに基づいてエネルギーが最小となるように計算する。橋本らは、高村らの手法を参考にしており、高村らはスピンモデルに基づいて単語の感情極性を計算した [Takamura 05a, Takamura 05a, Takamura 06]。ここで得られた単語の感情極性は『単語感情極性対応表』²¹として日本語と英語のものが公開されている。

²¹http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html

表 C.10: why 型質問応答の例

質問	回答候補	スコア
なぜ地球温暖化は進むのか？	二酸化炭素の温室効果は、同じ体積あたりではメタンやフロンにくらべ小さいものの、排出量が莫大であるため	0.873
	人口の10倍以上家畜が多いため	0.806
	有機化合物以外の物質を研究する化学は無機化学の範疇に含まれるため	0.384
何で原子力発電に反対なの？	希望が叶い大学入学後は原子力工学を専攻したが、その過程で現代の原子力工学における放射線被害の実態を知ったため	0.945
	原子力発電所の反対派については、ファッション、イデオロギー、信仰であり、反対派の多くが原子力発電所についての知識はなく、たとえ知識があってもイデオロギーから反対し、金儲けや売名のため	0.644
	社会党及び双葉地方原発反対同盟は福島第二立地点での反対運動に対して「支援」以上のことを出来なかったため	0.640
どうして人工知能は人間を超えられないのか？	「人工知能が人間知能を越えること」すなわち人間の脳は高速処理や大量処理には向いていないため	0.668
	知能指数のような人間向けの知能尺度を機械の知能にそのまま当てはめるのは簡単ではないため	0.657
	クレバーボットは、人間と会話するため	0.648

C.5 定義型質問応答

定義型質問応答は、「～とは？」のような定義を尋ねる質問に答えるものである。ノンファクトイド型の中でも、早くから研究がなされた分野である。定義型質問応答に関する研究は、未知語の定義獲得に関する研究と密接に関係しており、そこで得られた知見が生かされている。

代表的な定義型質問応答システムとしては、BBN テクノロジー社の BBN [Xu 03]、シンガポール大学の QUALIFIER [Yang 03]、復旦大学の FDUQA [Zhou 06]、Blair-Goldensohn らの DefScriber [Goldensohn 03]、IBM の PIQUANT [Carroll 05] などがある。

BBN は、定義型質問応答の最初の評価ワークショップである TREC-2003 においてベースラインとして BBN テクノロジー社が開発したシステムである [Xu 03]。単純な文書検索に基づくシステムであったにも関わらず、TREC-2003 では最も高性能な性能を示した。

QUALIFIER (QUestion Answering by LexIcal FabrIc and External Resources) は、シンガポール大学が開発した定義型質問応答システムである [Yang 03]。外部リソースとして Web 検索のスニペットを用いており、Wikipedia などの人手で作られた定義サイトを用いていないのが特徴である。

FDUQA は、復旦大学が開発した定義型質問応答システムである [Zhou 06]。外部リソースから得られる定義を特徴ベクトルに変換し、それを回答候補評価に用いている。

DefScriber は、質問応答システムではなく、定義作成システムである [Goldensohn 03]。人間が読みやすい定義を自動で作成することが目的である。定義文を 8 タイプ（属、種、類義語、部分、理由、効果、歴史、語源）に分類し、定義として特に重要とされる文を、機械学習と構文パターンのマッチングによって抽出する。

PIQUANT は、IBM が開発した定義型質問応答システムである [Carroll 05]。通常のアプローチとは異なり、定義型質問を複数のサブ質問に分割し、各サブ質問に対する回答を統合して最終的な回答を得るという独自の手法を採用している。

定義の自動獲得に関しても様々な研究が存在する。Coates-Stephens は、FUNES (Figuring-out Unknown Nouns from English Sentences) と呼ばれる Prolog ベースの定義獲得システムを開発した [Coates-Stephens 91]。「人間にとって未知の単語を含む文章には、書き手による説明文が存在するはずだ」と仮定し、未知の固有名詞を発見した際に、その文の構造と周辺情報から、固有名詞に関する name-frame と呼ばれるテーブルを埋めるようにして定義を獲得する。

Radev らは、PROFILE というシステムを開発し、新聞記事データから固有名詞を抽出し、これらを定義する名詞句を正規表現のパターンで獲得した [Radev 97]。西野らは、「A とは B」のパターンを用いて定義文を獲得する手法を提案した [西野 99]。「とは」には定義を表す場合と主題化を表す場合がある。そこで、頻出単語をもつものを除外するというアプローチをとった。

Joho らは、大規模なテキストデータの特性を利用し、query noun を含む文を全て抽出し、各文について KPW (key phrase weight), WC (word count), SN (sentence number) を計算し、これらの重み付け和で求まるスコアでランキングし、定義を獲得する手法を提案した [Joho 99].

また、近年では、機械学習により定義かどうかを識別する研究が多い [Degorski 08, Espinosa-Anke 13, Espinosa-Anke 14, Miliaraki 04, Xu 05, Yan 12]. 例えば、Yan らは、「～とは」のパターンの文に関して定義文かどうかを SVM で識別している [Yan 12].

さらに、獲得した定義文を分類・編集することで自動的に用語辞書を作成する研究も行われている [Blair-Goldensohn 03a, Blair-Goldensohn 03b, Blair-Goldensohn 04, 藤井 02, 桜井 02, 土田 04].

本研究では、人手で用語と定義表現にタグ付けしたデータを CRF で学習し、チャンキングを行うことで定義候補文を獲得する。そして、候補文に定義らしさをスコア付けして Ranking SVM で学習し適用することでランキングを行う。

また、対話での応用を考慮に入れ、定義文の獲得と同時に説明文の獲得も行う。定義型として獲得した表現は、定義を問う質問に対する応答としては有効であるが、対話システムが主体的に発話を生成する際には、活用しにくい。そのため、定義を述べるまでもないトピックに対して説明しようとした際に、文脈に応じてシステムが説明を行えるような知識を保持しておくことは有益である。

C.5.1 CRF によるチャンキング

まず、人手で用語と定義部および説明部に対してタグ付けを行う。

(例) : $\langle d \rangle$ 英語や日本語や中国語のように自然発生的に生まれた言語 $\langle /d \rangle$ を $\langle l \rangle$ 自然言語 $\langle /l \rangle$ と呼び、これに対して $\langle d \rangle$ 人為的に創作された言語 $\langle /d \rangle$ を $\langle l \rangle$ 人工言語 $\langle /l \rangle$ と呼ぶ。

(例) : $\langle r \rangle$ 自然言語 $\langle /r \rangle$ は $\langle e \rangle$ 母語として使用する人々の存在を前提として存在しているため、民族の滅亡や他言語による吸収によって使用されなくなることがある $\langle /e \rangle$ 。

そして、タグ付けされたデータを CRF で学習し、チャンキングを行う。なお、素性は why 型で説明した形態素素性を使用する。

表 C.11: 定義文／説明文獲得のための IOB ラベル

B-L	左側に定義/説明がある用語の開始
I-L	左側に定義/説明がある用語の内側
B-R	右側に定義/説明がある用語の開始
I-R	右側に定義/説明がある用語の内側
B-D	定義部の開始
I-D	定義部の内側
B-E	説明部の開始
I-E	説明部の内側
O	外側

C.5.2 ランキング学習による定義文／説明文らしさの評価

ランキング学習は教師あり学習の枠組みで検索ランキングを最適化する技術である．ランキング学習の手法として，Pointwise, Pairwise, Listwise という 3 つの手法が存在する．本研究では Pairwise 手法の一つである Ranking SVM を使用する．まず，CRF で取り出した定義型／説明型のデータを統合して書き出し，定義文らしさと説明文らしさを以下の基準に従ってスコアを付与した（表 C.12, C.13）．今回は，あるトピックに対して定義文／説明文候補が合計 5 つ獲得されたもののみを学習データとしてスコア付けした．その結果，合計 1863 個の用語からなる学習データを構築した．そして，Ranking SVM の学習には SVM^{rank} を使用し，定義型／説明型それぞれの識別器を構築した．なお，カーネルとして線形カーネルを使用し，緩和パラメータの値は 1.0 に設定した．そして，学習した識別器を適用して全ての用語に対して定義文および説明文のランキングを行った．

定義型質問応答の例を表 C.14 に示す．ランキング学習の効果もあり，定義型識別器では定義文らしいものが上位に来ており，説明型識別器では説明文らしいものが上位に来ている．

現状の問題点として，多義性を考慮できていないこと，回答候補が多い用語（e.g. 日本）のランキング結果がそれほど良くないことが挙げられる．前者の問題は，対話システムで用いる場合，多義性は文脈を用いることで対処可能だと考えられる．後者の問題は，フィルタリング処理などを行い，できるだけ候補を絞り込む必要がある．

表 C.12: 定義型のスコア付け基準

スコア	基準	例
1	文法的/意味的に不適切	人間は、動物の
2	ドメインが限定的／冗長	人間は、争いを好む
3	類が非明示的	人間は、言葉でコミュニケーションする
4	種差が非明示的	人間は、動物（最近類）の一種である
5	最近類と種差が明示的	人間は、理性的な（種差）動物（最近類）である

表 C.13: 説明型のスコア付け基準

スコア	基準	例
1	文法的／意味的に不適切	人間は、動物の
2	ドメインが限定的／冗長	人間は、争いを好む
3	類で定義づけられている	人間は、理性的な（種差）動物（最近類）である
4	説明的だが説明不足	人間は、古来より人間自身を思考の対象としてきた
5	用語の性質を広く捉えた説明	人間は、文字や言語を抽象的なシンボルとして扱ったり、論理思考を行い、多様な事象に様々な解釈を行う

表 C.14: 定義型／説明型の質問応答の例

質問対象語	定義型回答候補	定義型適合度	説明型回答候補	説明型適合度
核酸	塩基と糖，リン酸からなるヌクレオチドがリン酸エステル結合で連なった生体高分子	0.718	1 8 6 8 年にフリードリッヒ・ミーシャーにより発見された	0.993
	直鎖状に多数接続してできたポリマー	0.181	直鎖状に多数接続してできたポリマー	0.389
	1 8 6 8 年にフリードリッヒ・ミーシャーにより発見された	0.105	塩基と糖，リン酸からなるヌクレオチドがリン酸エステル結合で連なった生体高分子	0.270
ザグレブ	クロアチア共和国の首都で同国最大の都市	0.514	2 0 1 0 年の観光客数は 6 0 万人以上で，2 0 1 1 年には 1 0 % 増加が見込まれているなど，増加傾向にある	1.246
	国際空港の他にルチェコ空港があり，スポーツ用の飛行機やクロアチア警察の特殊部隊，陸軍のヘリコプターが常駐する	0.291	数多くの博物館はザグレブとクロアチアの歴史，美術および文化を反映しているばかりでなく，ヨーロッパや世界のものも収めている	1.217
	2 0 の常設や季節的な劇場やステージ	0.205	クロアチアの 5 つの主要高速道路の要衝で，ここ数年来すべてのクロアチアの高速道路はザグレブを起終点としている	1.133

C.5.3 「とは」を手がかり表現とした定義獲得

「NP とは」で始まる文が必ずしも定義文とは限らないため、識別器で定義文か否かを識別する。まず、「NP とは」とのパターンマッチングで『日本語ウェブコーパス 2010』²²から 23,004,853 個の定義文候補を抽出した。Web コーパスから抽出した定義文候補からランダムに 4,000 個取り出しラベルを振り（正例:604 個，負例:3396 個），正例に Wikipedia において「NP とは」で定義づけられる定義文 3,000 個を追加したものをデータセットとする。

非定義文の特徴は文末と「NP とは」の直後に現れやすいことから，文末の 5 形態素の単語属性とその 2-gram, 3-gram, および，「NP とは」直後の 5 形態素の単語属性とその 2-gram, 3-gram を素性として利用した。ここで，単語属性とは形態素の {表層形, 基本形, 品詞大分類, 品詞細分類} を表す。識別器として RandomForest と線形カーネル SVM を使用し，10 分割交差検定で評価した。なお，RandomForest の決定木の本数は 200 個に設定し，乱雑さの評価基準としてジニ不純度を使用した。また，SVM のコストパラメータは 1.0 に設定した。

実験結果を表 C.15, C.16 に示す。定義を定義と判定する精度においては，RandomForest の方が良い結果を示したが，非定義を非定義と判定する精度も考慮した場合，平均精度は線形カーネル SVM の方が良いため，定義獲得では線形カーネル SVM を使用した。この識別器を Web コーパスから抽出した 23,004,853 個の定義文候補に対して適用することで，重複なしで 5,327,918 個の定義文を獲得した。なお，用語の異なり数は 860,809 個である。

表 C.15: 定義識別結果 (RandomForest)

	定義	非定義	精度	データ数
定義	3432	172	95.2%	3604
非定義	353	3043	89.6%	3396

表 C.16: 定義識別結果 (線形カーネル SVM)

	定義	非定義	精度	データ数
定義	3419	185	94.9%	3604
非定義	207	3189	93.9%	3396

²²<http://www.s-yata.jp/corpus/nwc2010/>

C.5.4 「類」の抽出

定義文はしばしば最近類 (genus proximum) と種差 (differentia specifica) によって定義付けられる。最近類とは対象を種として含む類 (上位概念) のことであり、種差とは対象を他の種から区別する特徴である。例えば、「人間とは理性的な動物である」という文において、「動物」が最近類であり、「理性的な」が種差である。

ここでは、前章で獲得した定義文から類を抽出することが目的である。類の抽出には識別モデルを使用する。チャンキングのラベルとして IOBES ラベルを用いた (表 C.17)。識別モデルとして CRF (Conditional Random Fields) と CNF (Conditional Neural Fields) を使用し、素性として形態素単位の素性または文字単位の素性を使用した。形態素素性には、前後 K 個の単語属性の N -gram ($K = 5, N = 1, 2, 3$) を用いた。ここで、単語属性とは {単語表層形, 品詞大分類, 品詞細分類, 対象単語が用語に含まれているかどうか} を表す。また、文字素性には、前後 K 個の文字属性の N -gram ($K = 5, N = 1, 2, 3$) を用いた。ここで、文字属性とは {文字の表層形, 品詞大分類の IOBES タグ, 品詞細分類の IOBES タグ, 字種 (ひらがな, カタカナ, 数字, アルファベット, 記号, その他), 対象文字が用語に含まれているかどうか} を表す。

各識別モデルを各素性で学習し、10 分割交差検定で評価した結果を表 C.18, C.19, C.20, C.21 に示す。このうち、F 値の平均が最も高かったのは、文字素性を使用したときの CRF の結果である。そこで、この CRF を獲得した定義文に対して適用し、各用語の類を抽出した。その結果、例えば「急性肝炎」の類として {肝障害, 肝臓病, 肝臓障害, 肝機能障害, 肝炎, 病気, ...} などが抽出された。

表 C.17: ラベルの種類

I	チャンクの内側
O	チャンクの外側
B	チャンクの開始
E	チャンクの終わり
S	単体でチャンクを構成

表 C.18: CRF による類の抽出結果
(形態素素性)

	精度	再現率	F 値	ラベル数
I	92.3	54.6	68.6	245
O	97.7	99.0	98.4	18965
B	86.7	76.5	81.3	485
E	88.9	78.4	83.3	485
S	83.1	79.7	81.4	810
平均	89.7	77.6	82.6	

表 C.19: CRF による類の抽出結果
(文字素性)

	精度	再現率	F 値	ラベル数
I	95.9	76.5	85.1	1842
O	97.5	99.4	98.5	34113
B	94.2	82.4	87.9	1183
E	94.3	83.2	88.4	1183
S	62.5	83.3	71.4	112
平均	88.9	85.0	86.3	

表 C.20: CNF による類の抽出結果
(形態素素性)

	精度	再現率	F 値	ラベル数
I	68.1	87.0	75.5	245
O	99.0	98.1	98.5	18965
B	78.3	85.1	81.4	485
E	80.9	87.8	84.1	485
S	80.8	86.5	83.5	810
平均	81.4	88.9	84.6	

表 C.21: CNF による類の抽出結果
(文字素性)

	精度	再現率	F 値	ラベル数
I	71.2	86.7	77.9	1842
O	98.6	97.3	98.0	34113
B	79.3	83.6	81.4	1183
E	83.0	87.6	85.2	1183
S	73.6	89.3	79.7	112
平均	81.1	88.9	84.4	

C.6 ニューラルネットワークを用いた回答段落選択

QANet [Yu 18] の構造を参考に、文書から回答を含む段落を選択するモデルを構築した (図 C.3)。QANet は、正解の範囲 (単語の開始位置と終了位置) を出力するモデルなのに対し、このモデルは段落が回答を含むかどうかを系列ラベリングの要領で判定する。

モデルは質問の単語系列とニュース記事の単語系列をエンコードした結果に対して Context-Query Attention [Yu 18] を計算する。次に段落ごとに self-attention [Lin 17] を計算して得られた段落の埋め込み表現を双方向 LSTM に与える。その結果と段落における質問単語集合の被覆率を結合したものを出力層に与え、各段落が回答を含む確率を計算する。

Text Encoder を図 C.4 に示す． Embedding Encoder は，QANet の Encoder Block と同じである（図 C.5）． Position Encoding [Vaswani 17] で単語の出現位置の埋め込み表現を得る．そして，畳み込み層により局所的な特徴を，self-attention 層 [Vaswani 17] により大域的な特徴を捉えるように計算する．なお，各中間層の出力は Layer Normalization [Ba 16] により正規化することで学習の効率化を図っている．

Input Embedding は，補助情報を加えている以外は QANet の Input Embedding の構造と同じである（図 C.6）． word2vec [Mikolov 13a, Mikolov 13b, Mikolov 13c, 西尾 14] で単語と文字の分散表現を事前に計算しておく．文字の分散表現はさらに畳み込みで単語レベルのベクトル表現に変換し [Kim 14]，単語の分散表現と補助情報を組み合わせて Highway Network [Srivastava 15] に与える．

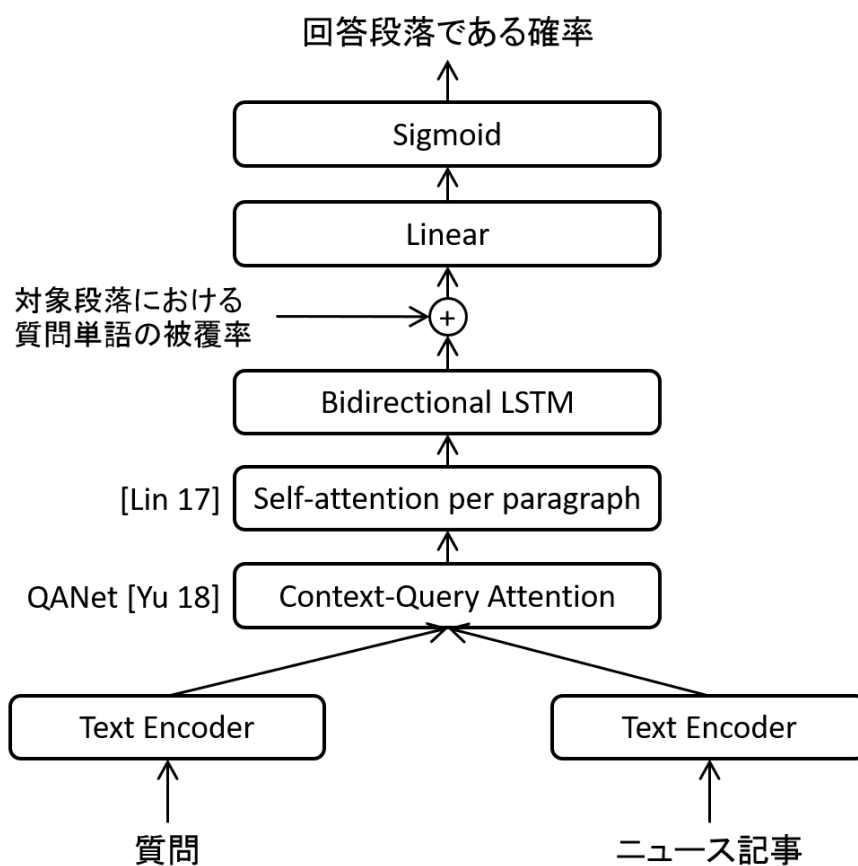


図 C.3: 回答段落選択モデル

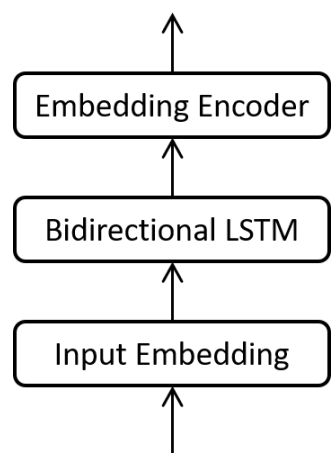


図 C.4: Text Encoder

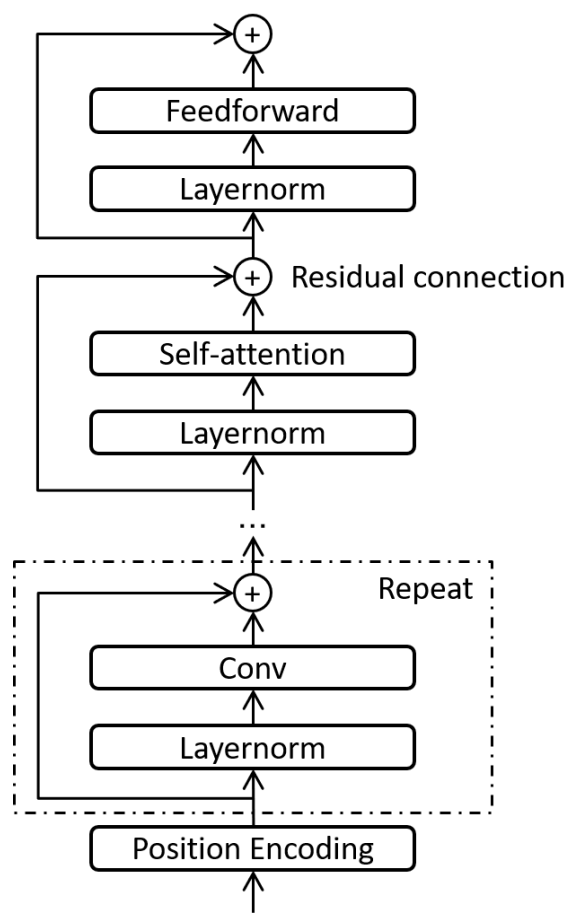


図 C.5: Embedding Encoder

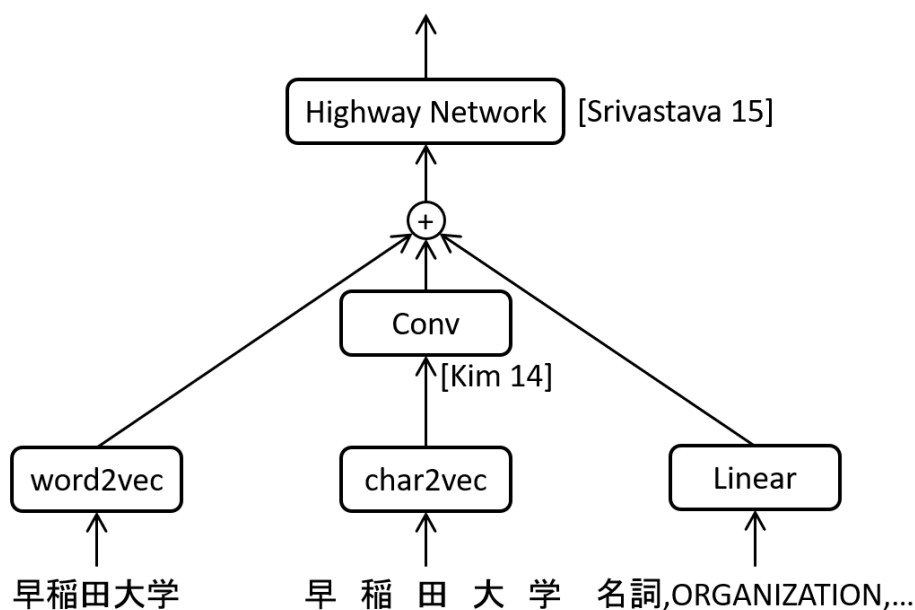


図 C.6: Input Embeddings

C.6.1 QA データの収集

Yahoo!ニュース²³の「ココがポイント」に書かれている QA データを使用した。例えば、「日本最古の老舗企業は？」という質問に対して、「宮大工の流れを組む社寺建築業者「金剛組」。創業は西暦 578 年、創業 1400 年を超えており、世界一歴史がある企業。」という回答が付与されている。回答には元記事へのリンクが貼られており、そこに書かれている内容の要約が回答となっている。そこで、回答を句点「。」で区切った各文と、リンク先の記事の段落との類似度を計算し、最も類似度が高い段落を正解とした。回答が複数文から構成されている場合、複数段落が正解になりうる。今回の実験では、876 記事を訓練セット、100 記事をテストセットとした。

C.6.2 実験設定

確率値でランキングし、トップの段落を回答とする。ただし、回答が N 段落から構成されている場合、上位 N 個を回答とする。ベースラインとして、ランダムに段落を選択したモデル (Random) と、段落における質問単語集合の被覆率でランキングするモデル (Coverage) と比較した。

²³<https://news.yahoo.co.jp/>

C.6.3 実験結果

実験結果を図 C.7 に示す．この結果から，Random や Coverage よりも提案モデルの方が正解率が高いことが分かった．

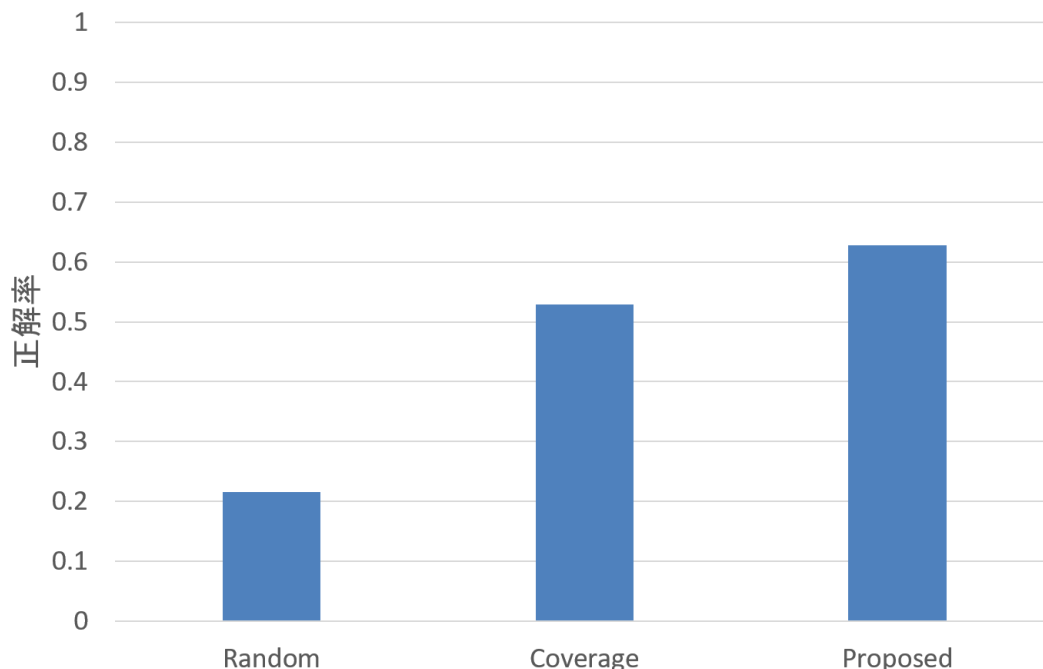


図 C.7: 実験結果

C.7 まとめと今後の課題

オープンドメイン質問応答のうち，ファクトイド型およびノンファクトイド型の why 型と定義型についてシステムを構築した．現状のシステムは，性能が高いとは言えない．今後は，ニュースを伝えるというタスクに焦点を当てた質問応答の手法について検討する．

ニューラルネットワークを用いた回答段落選択では，単語の分散表現を得るのに word2vec を利用したが，それ以外にも GloVe [Pennington 14] や fastText [Bojanowski 17, Joulin 17] などがある．また，文書の分散表現を計算する手法として，paragraph2vec [Le 14] や Skip-thought [Kiros 15] などがある．今後は，分散表現の計算方法を変えてみたり，Text Encoder として BERT [Devlin 18] を使うなどして正解率の向上を狙う．また，今回は記事内に正解があることを前提としてモデル化したが，実際の QA では，記事中に回答がない場合もある．今後は，先行研究 [Hu 18, Clark 18, Levy 17] を参考にしながら，回答がないことを理解する仕組みについても検討したい．

付 録 D 要約コーパスの分析

2.3.2 節で説明した日経コーパスの統計の詳細を記す。

D.1 日経コーパスの設計

日経新聞のテクノロジー系のニュース記事 100 個について 3 人の作業者に要約を作成させた。1 記事あたりの平均文数は 35 文である。要約作業は、重要文抽出、整列、文圧縮の 3 つの作業からなる。

重要文抽出は、ニュース記事の中から重要な文を数文抽出する作業である。ここで、重要文を 2 種類定義した。一つは、見出し的な内容・核となる情報を含む数文を抽出したものである。ここで抽出した文をタイプ a とし、タイプ a の文集合からなる要約を要約 A とする。もう一つは、要約 A の内容を補足するような文を数文抽出し、要約 A に加えたものである。ここで抽出した文をタイプ b とし、タイプ a とタイプ b の文集合からなる要約を要約 B とする。作業には、タイプ a の文を 3 から 6 文、タイプ b の文を 2 から 5 文の範囲で抽出するように指示した。

整列は、抽出した重要文を会話で伝えるのに適した順番に並び替える作業である。作業には、落ちや教訓、今後の展望など会話の締めとなる情報が最後にくるような構成にするように指示した。

文圧縮は、抽出した重要文を会話で伝えるのに適した情報量に削減する作業である。ここでは、各文について会話でこう話したいという口語表現を最初に考えさせ、その口語表現を実現するのに必要な情報を文節単位で選択させた。

D.2 日経コーパスの作業結果の分析

作業者ごとの 1 トピックあたりの平均選択文数を図 D.1 に示す。重要文 a の選択文数ごとのヒストグラムを図 D.2 に示す。重要文 b の選択文数ごとのヒストグラムを図 D.3 に示す。重要文 a と重要文 b を合わせたときの選択文数ごとのヒストグラムを図 D.3 に示す。これらから、重要文 a と重要文 b が 1 トピックあたり 3, 4 文選ばれていることが分かる。

各作業者が 1 文目を含めたかどうかを要約 A と要約 B について調べた結果を図 D.5 に示す。要約 A と要約 B が同数なのは、1 文目が重要文 a として選ばれたためである。より詳細に選択

された文の位置ごとの頻度を調べた結果を図 D.6 と図 D.7 に示す。図 D.6 から重要文 a は記事の前半から多く選ばれる傾向にあることが分かり、図 D.7 から重要文 b は記事の中央から多く選ばれる傾向にあることが分かる。

2 組の作業者の組み合わせに対して Cohen's Kappa を計算し、作業者全員に対して Fleiss' Kappa を計算した結果を図 D.8 に示す。作業者の一方をシステム出力、他方を正解として重要文の選択精度を計算した結果を図 D.9 に、ROUGE-1 を計算した結果を図 D.10 に示す。これらの結果から、重要文 a の一致率に対して重要文 b の一致率が低いことが分かる。これは、図 D.6 と図 D.7 の結果から分かる通り、重要文 a は作業者に依らず記事の前半から選ばれるのに対し、重要文 b は作業者によって様々な場所から選ばれる傾向にあるためだと考えられる。

要約 A と要約 B に関する各作業者の文圧縮結果の平均要約率を表 D.1（要約 A）と表 D.2（要約 B）に示す。要約率は以下の式で計算した。

$$\text{要約率} = \frac{\text{文圧縮後の文節数}}{\text{元の文の文節数}} \quad (\text{D.1})$$

この結果から、作業者 2 の結果が他の作業者に比べて、要約率が低い（圧縮率が高い）ことが分かる。また、作業者間で選ばれた同じ文に対して文圧縮結果（文節選択）の一致率を計算した結果を図 D.11（要約 A）と図 D.12（要約 B）に示す。Accuracy はおよそ 70% 程度であった。

整列結果が記事の出現順と同じであったかどうかを確認した結果を図 D.13 に示す。要約 A に関しては過半数において記事の順番と同じであるのに対し、要約 B では重要文 b の順番が入れ替えられるケースが多かった。

日経コーパスの各記事の各段落が何文で構成されているか頻度を計算した結果を図 D.14 に示す。この結果から、1 段落あたり 2, 3 文から構成されていることが分かる。

重要文を含む段落を重要段落と呼ぶことにする。このとき、重要文抽出作業の結果、いくつの段落が重要段落として選ばれたことになるか計算した結果を図 D.15 に示す。この結果から、要約 A は 2 段落程度、要約 B は 4 段落程度選ばれていることが分かる。

重要段落として選ばれた段落の数のヒストグラムを図 D.16（要約 A）と図 D.17（要約 B）に示す。この結果から、要約 A については、作業者 3 はほとんど 1 段落からしか文を選んでいないことが分かる。一方、作業者 2 は 1 から 2 段落程度、作業者 1 は 1 から 3 段落程度から文を選んでいる。要約 B については、作業者 1 は 3 段落がピーク、作業者 2 は 5 段落がピーク、作業者 3 は 4 段落がピークになっており、重要文 b として様々な段落から文が選ばれていることが推察される。

重要段落の位置ごとの頻度を計算した結果を図 D.18（要約 A）と図 D.18（要約 B）に示す。この結果から、要約 A は先頭段落の文から構成される傾向にあるのに対し、要約 B は重要文 b

が記事の中央から選ばれる傾向にあるため（図 D.7），中央の重要段落頻度が要約 A よりも高くなっていることが分かる．

重要段落が包含する重要文の数の頻度を計算した結果を図 D.20（要約 A）と図 D.21（要約 B）に示す．要約 A については，作業者 1 と作業者 2 は過半数が 1 段落あたり 1 文なのに対し，作業者 3 は 1 段落あたり 3 文程度選んでいることが分かる．要約 B については，重要文 b が様々な個所から選ばれることから，全作業者について過半数が 1 段落あたり 1 文の頻度が高くなっていることが分かる．

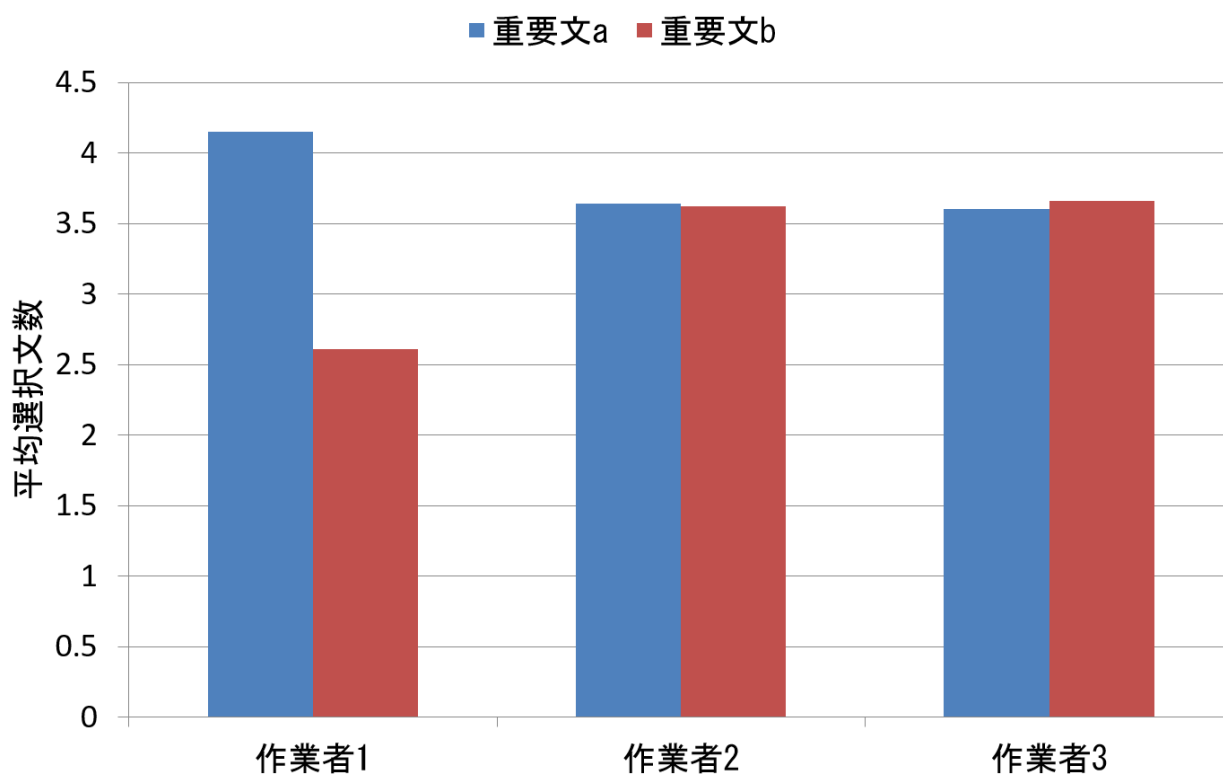


図 D.1: 重要文抽出作業結果：1 トピックあたりの平均選択文数

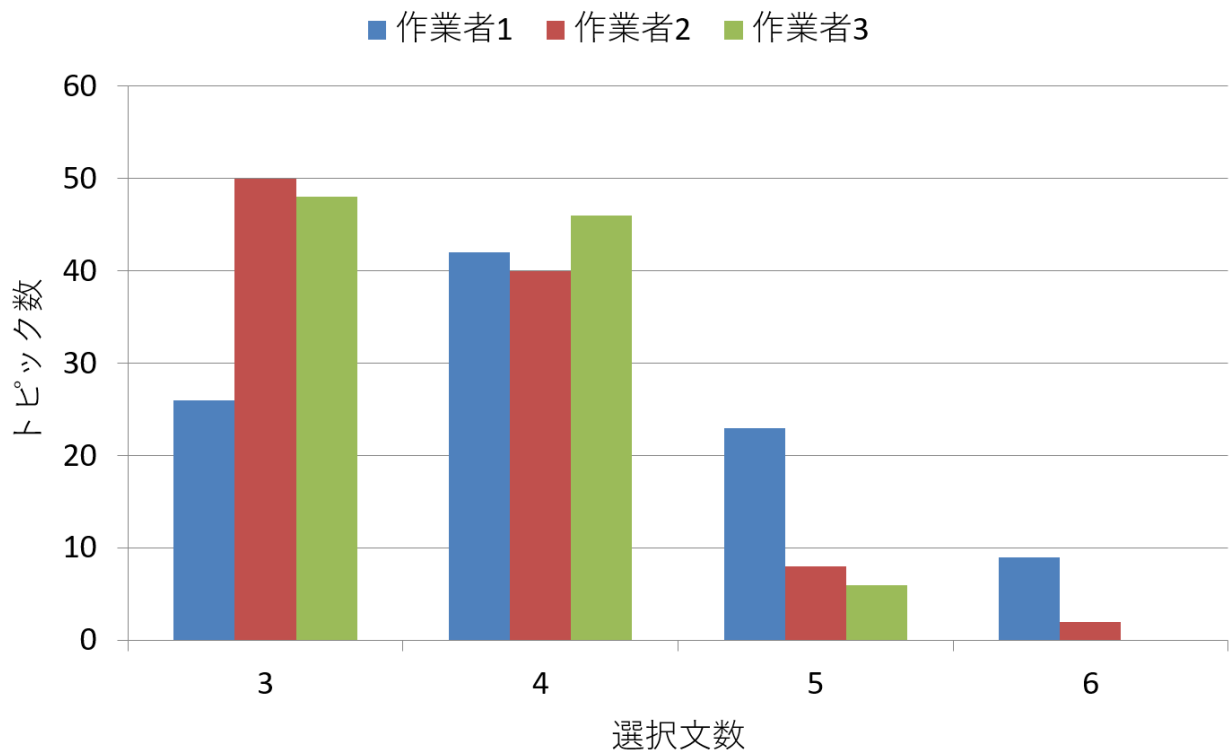


図 D.2: 重要文抽出作業結果：選択文数のヒストグラム（重要文 a）

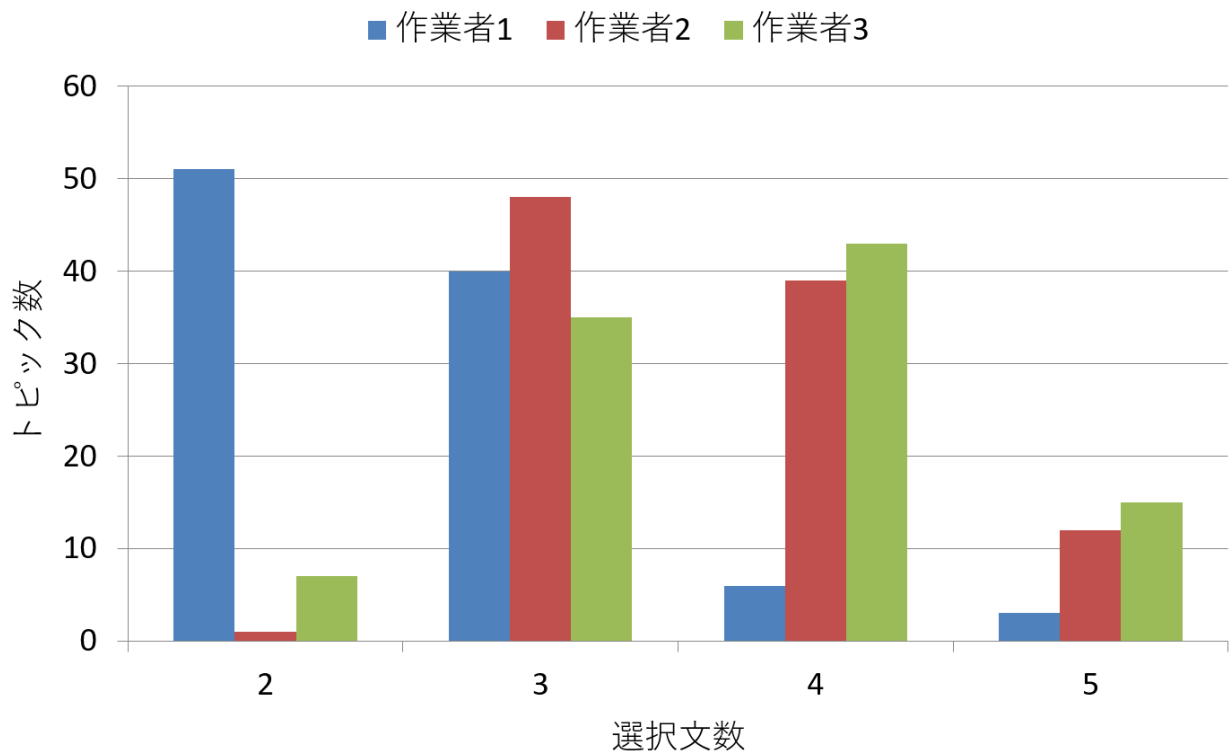


図 D.3: 重要文抽出作業結果：選択文数のヒストグラム（重要文 b）

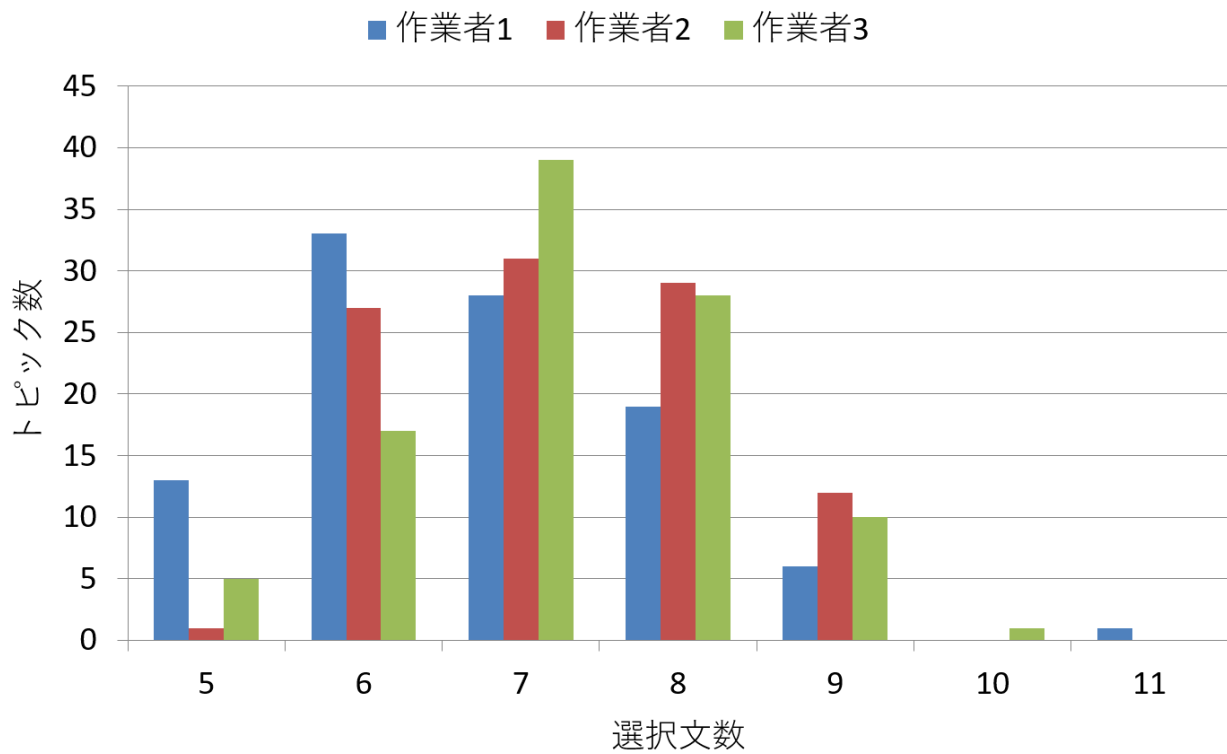


図 D.4: 重要文抽出作業結果：選択文数のヒストグラム（重要文 a+重要文 b）

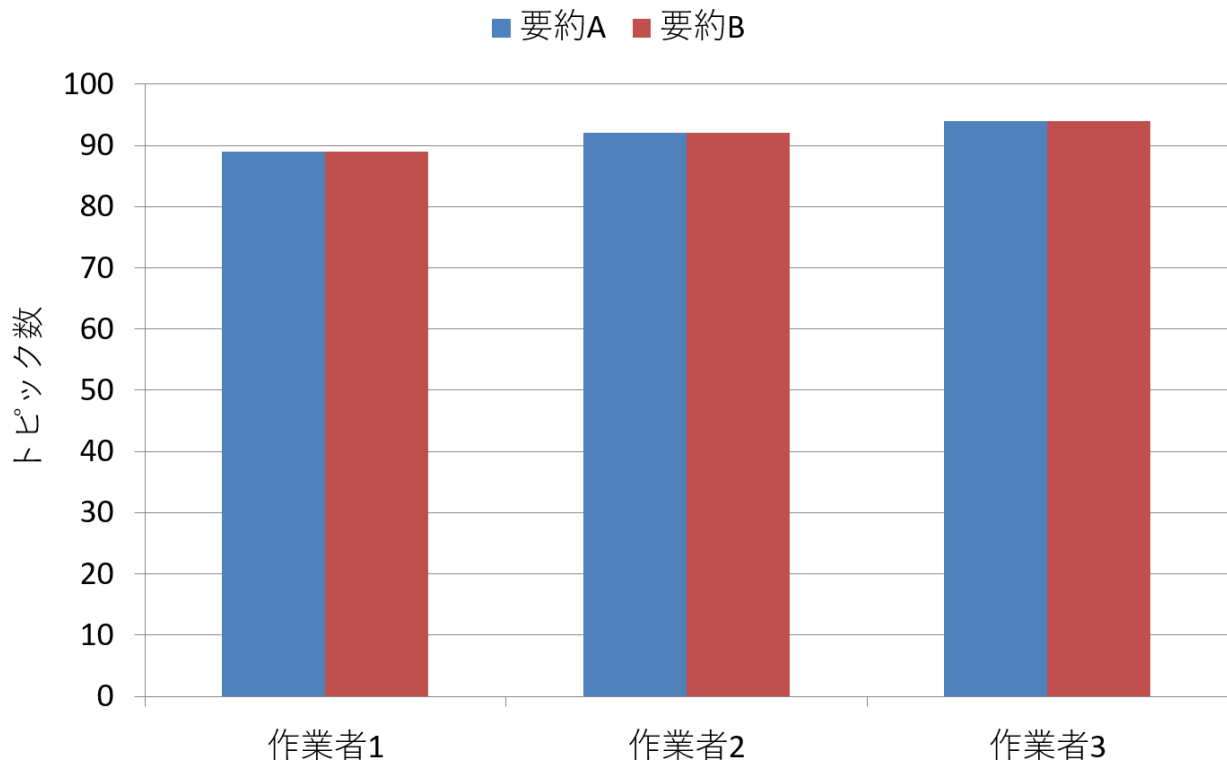


図 D.5: 重要文抽出作業結果：1文目が選ばれたものの数

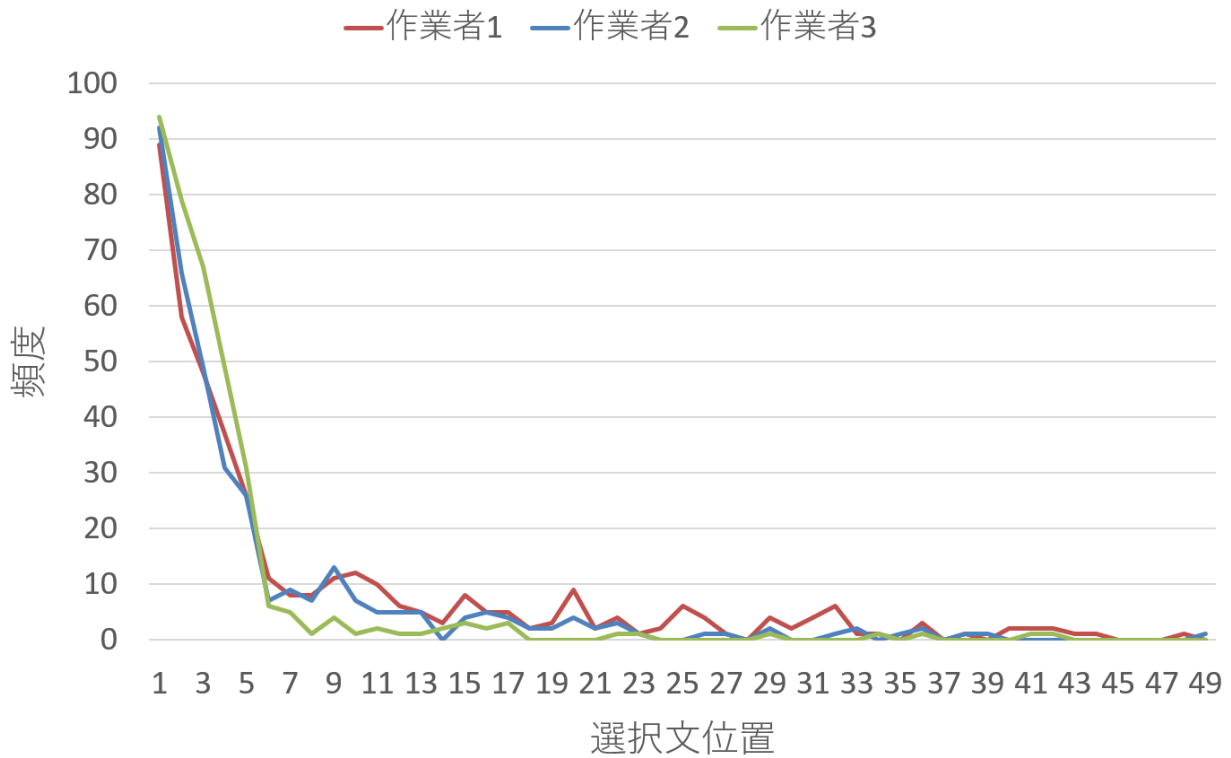


図 D.6: 重要文抽出作業結果：選択文位置のヒストグラム（重要文 a）

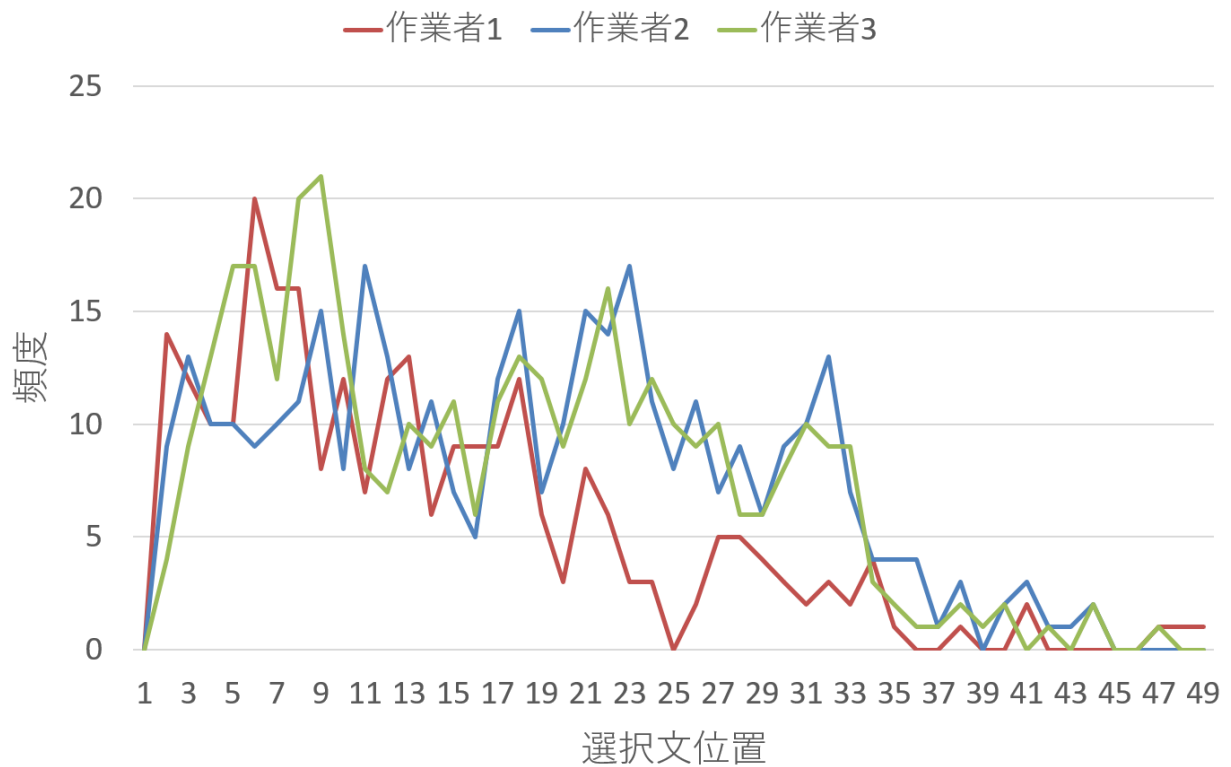


図 D.7: 重要文抽出作業結果：選択文位置のヒストグラム（重要文 b）

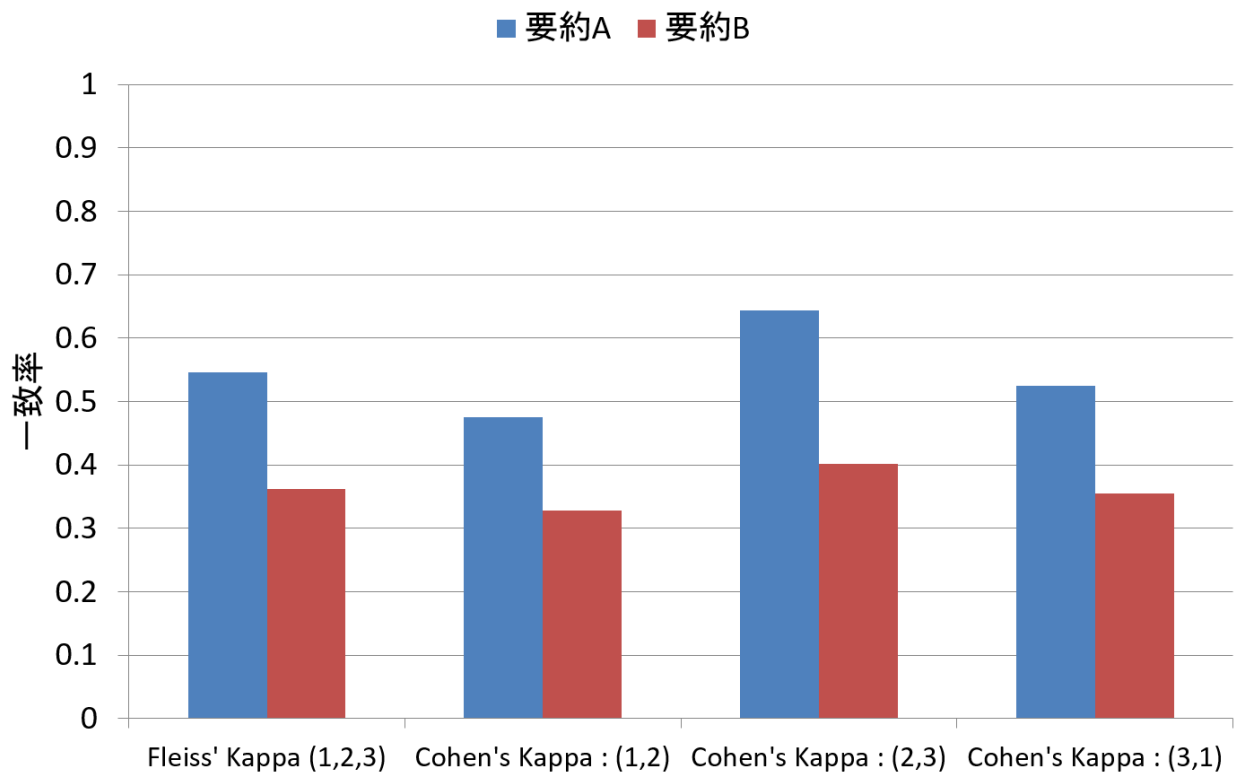


図 D.8: 重要文抽出作業結果：1 文目を含めたものの頻度

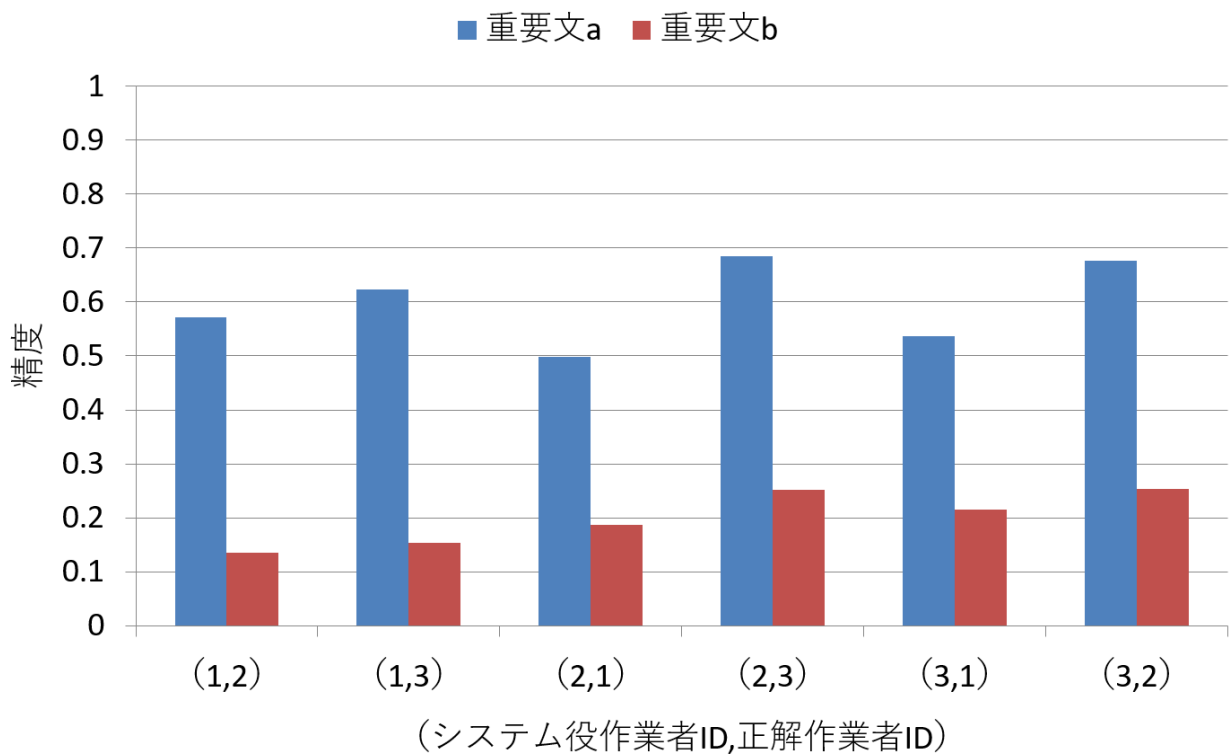


図 D.9: 重要文抽出作業結果：選択精度

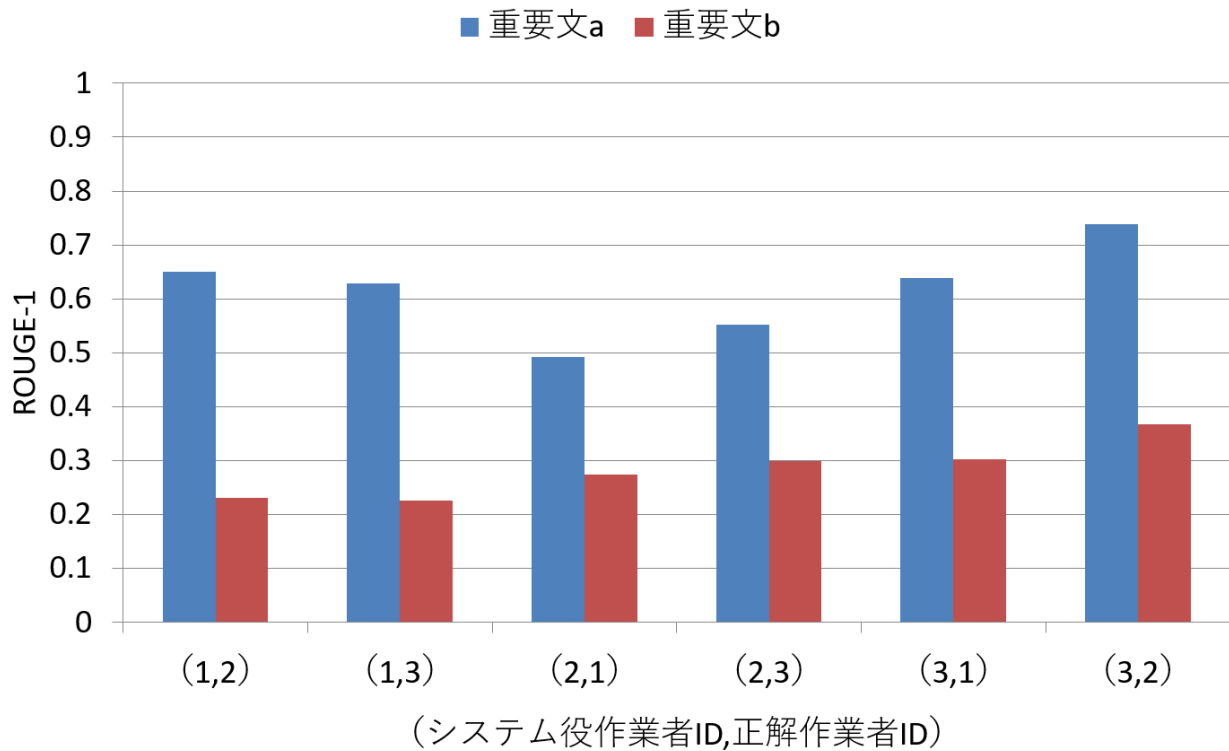


図 D.10: 重要文抽出作業結果：ROUGE-1

表 D.1: 文圧縮作業結果：要約 A に関する各作業者の平均要約率（％）

	選択された文全体の平均要約率	文圧縮が行われた文の平均要約率	選択された文のうち、文圧縮が行われた文の割合
作業者 1	77.9	70.9	75.9 (315 / 415)
作業者 2	63.5	59.0	90.0 (322 / 362)
作業者 3	84.1	73.2	59.3 (212 / 358)

表 D.2: 文圧縮作業結果：要約 B に関する各作業者の平均要約率（％）

	選択された文全体の平均要約率	文圧縮が行われた文の平均要約率	選択された文のうち、文圧縮が行われた文の割合
作業者 1	78.1	71.3	76.5 (517 / 676)
作業者 2	65.5	60.9	88.0 (639 / 726)
作業者 3	83.2	71.7	59.4 (431 / 726)

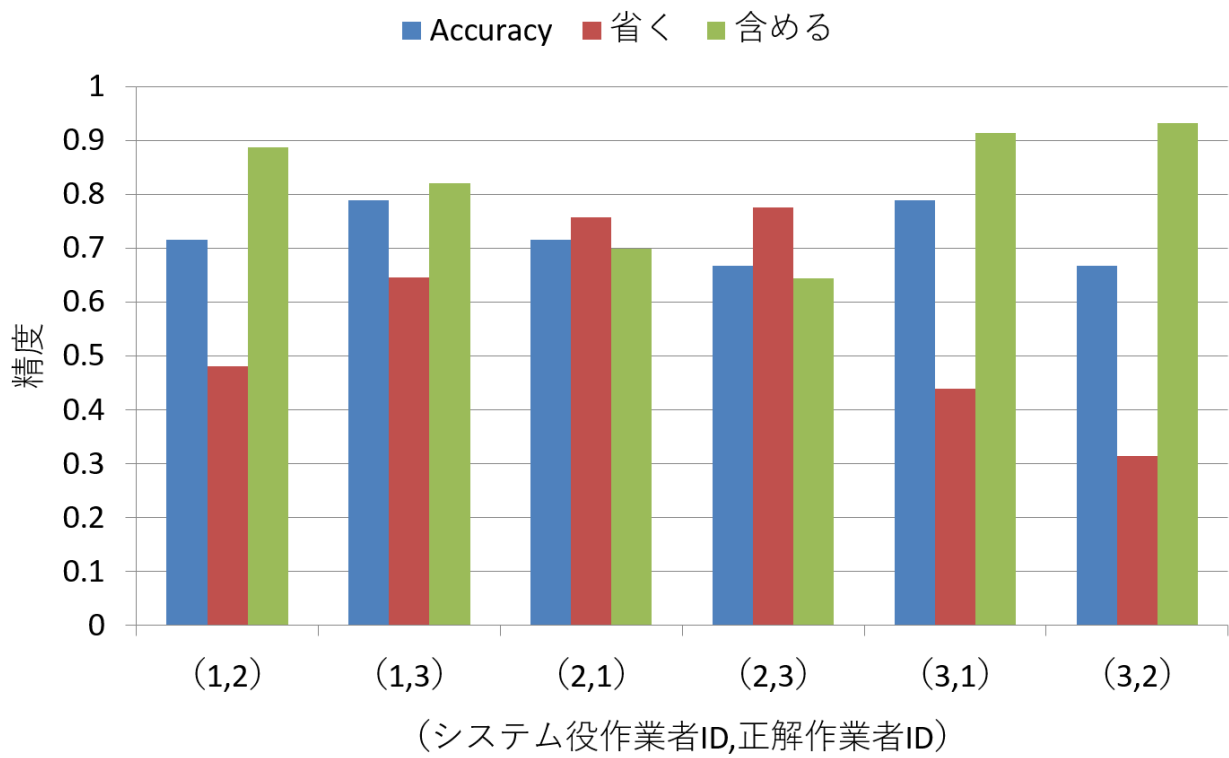


図 D.11: 文圧縮作業結果：文節の一致率（要約 A）

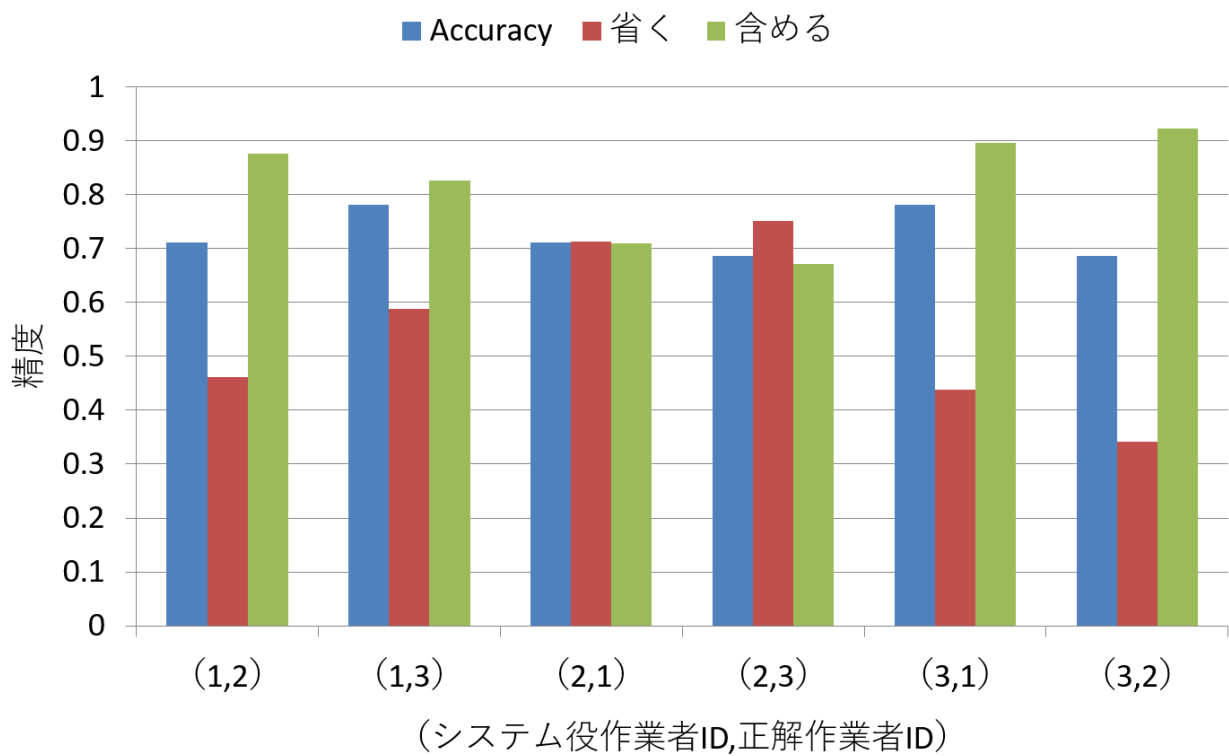


図 D.12: 文圧縮作業結果：文節の一致率（要約 B）

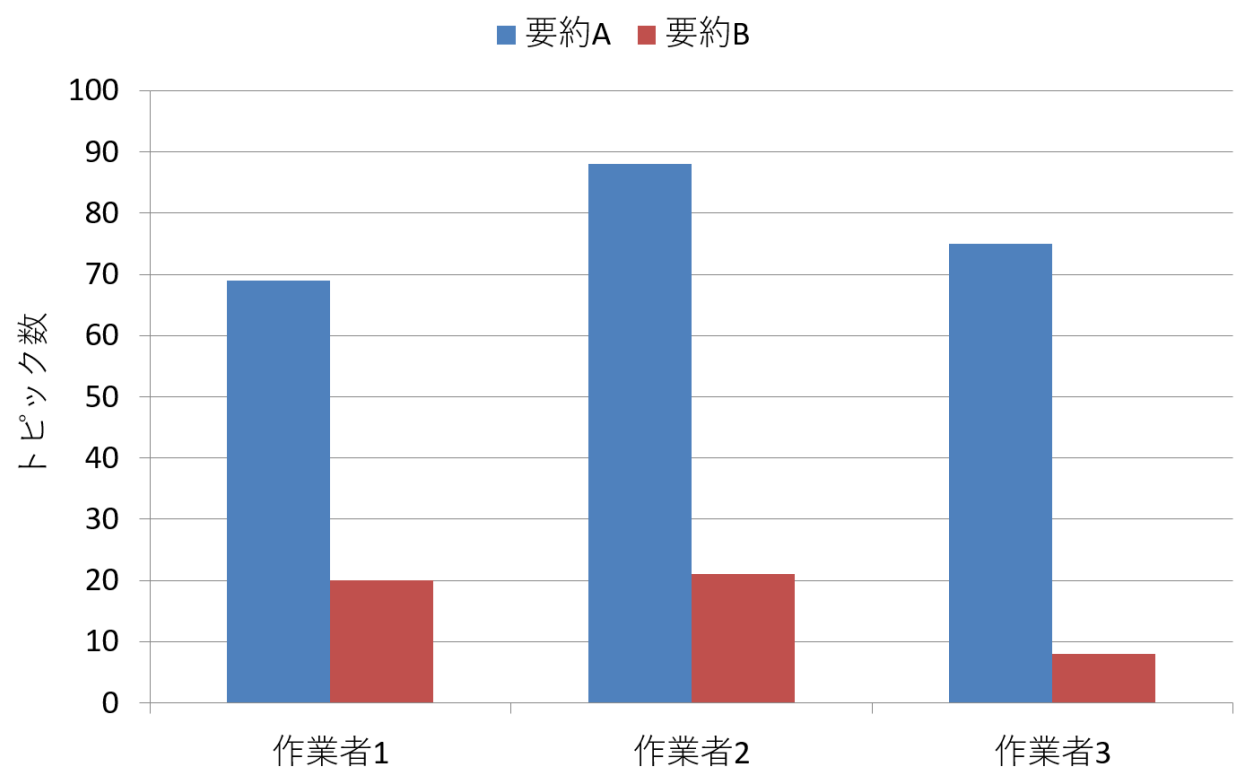


図 D.13: 整列作業結果：記事の出現順と一致していたものの数

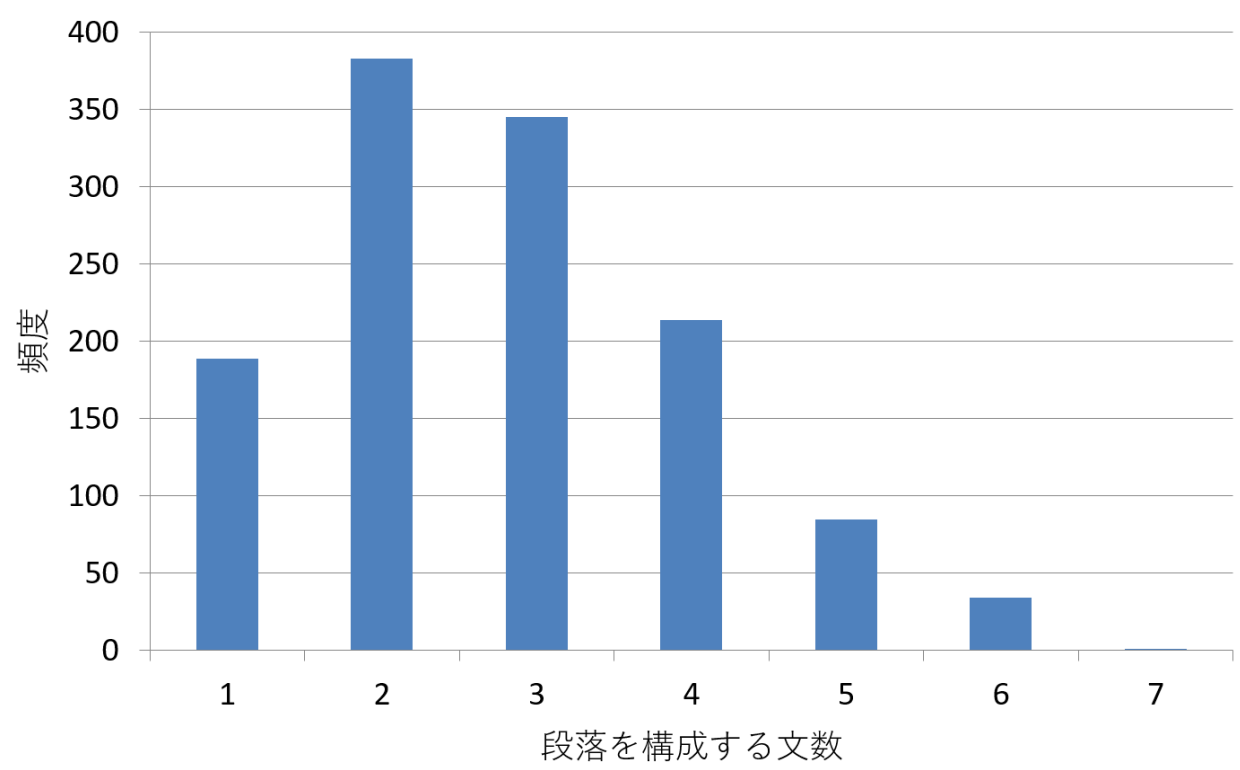


図 D.14: コーパスの記事の各段落を構成する文の数のヒストグラム

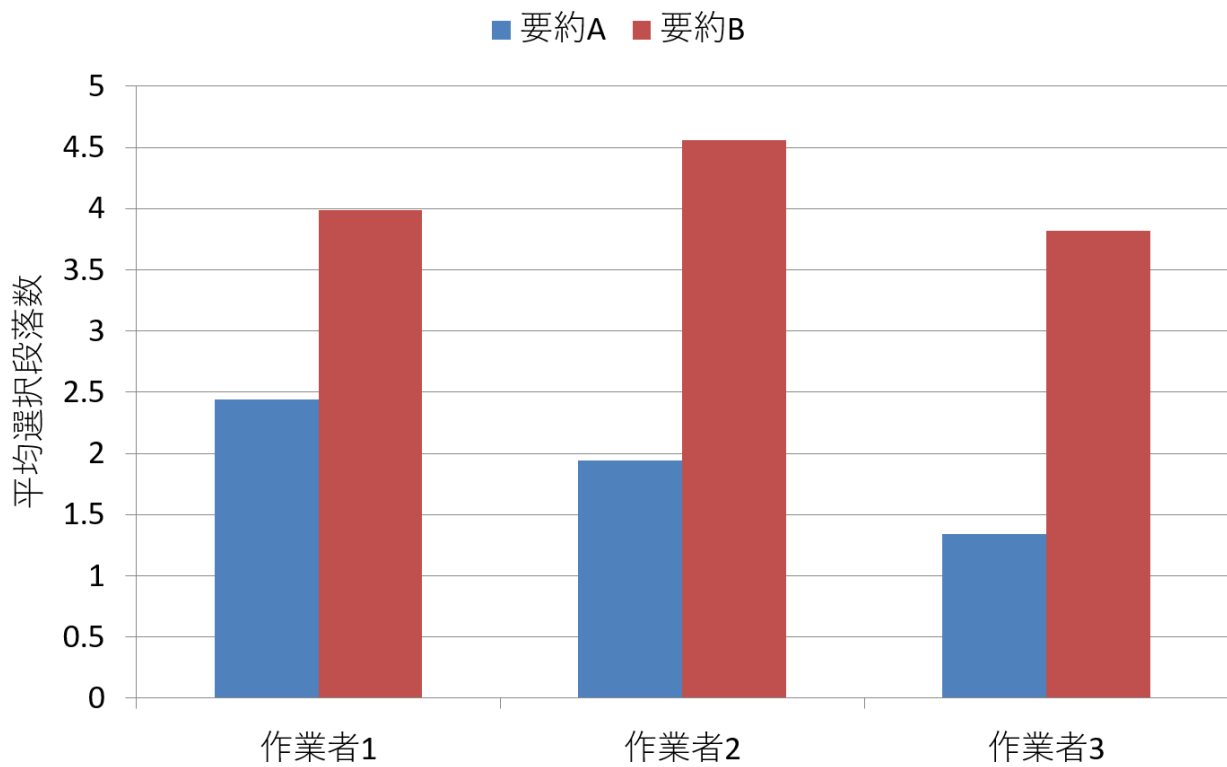


図 D.15: 平均重要段落数

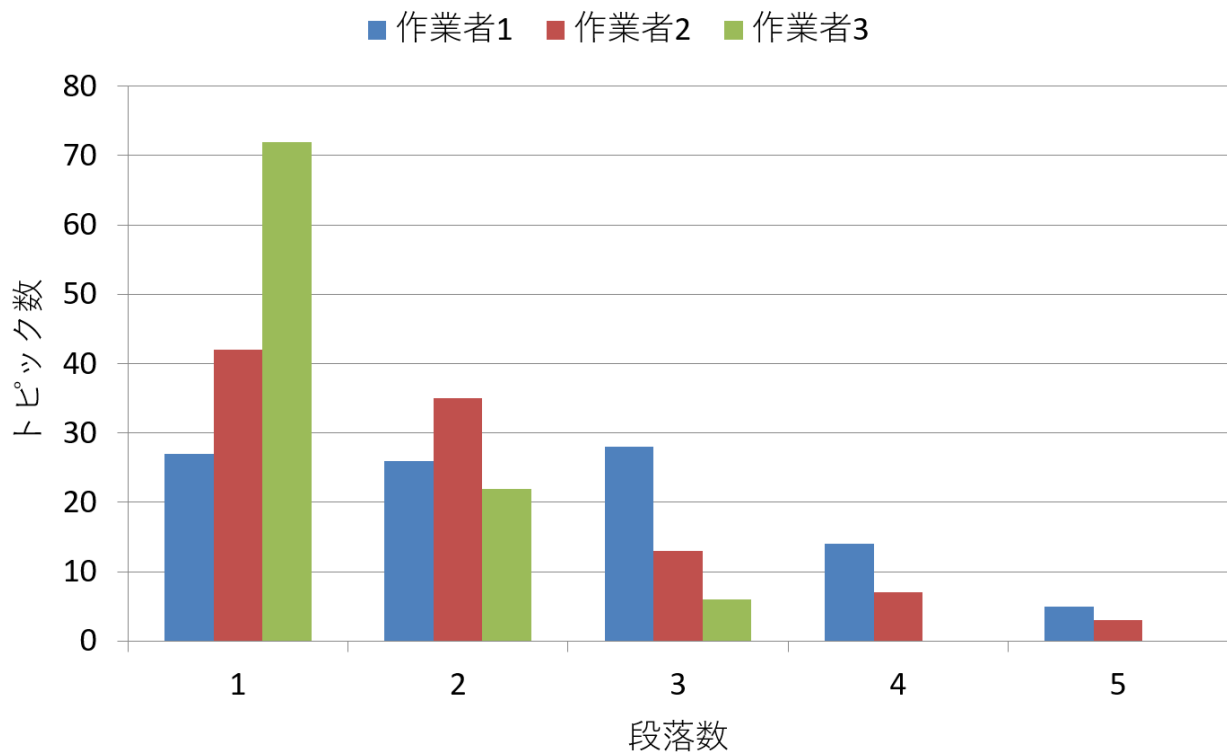


図 D.16: 重要段落として選ばれた段落の数のヒストグラム (要約 A)

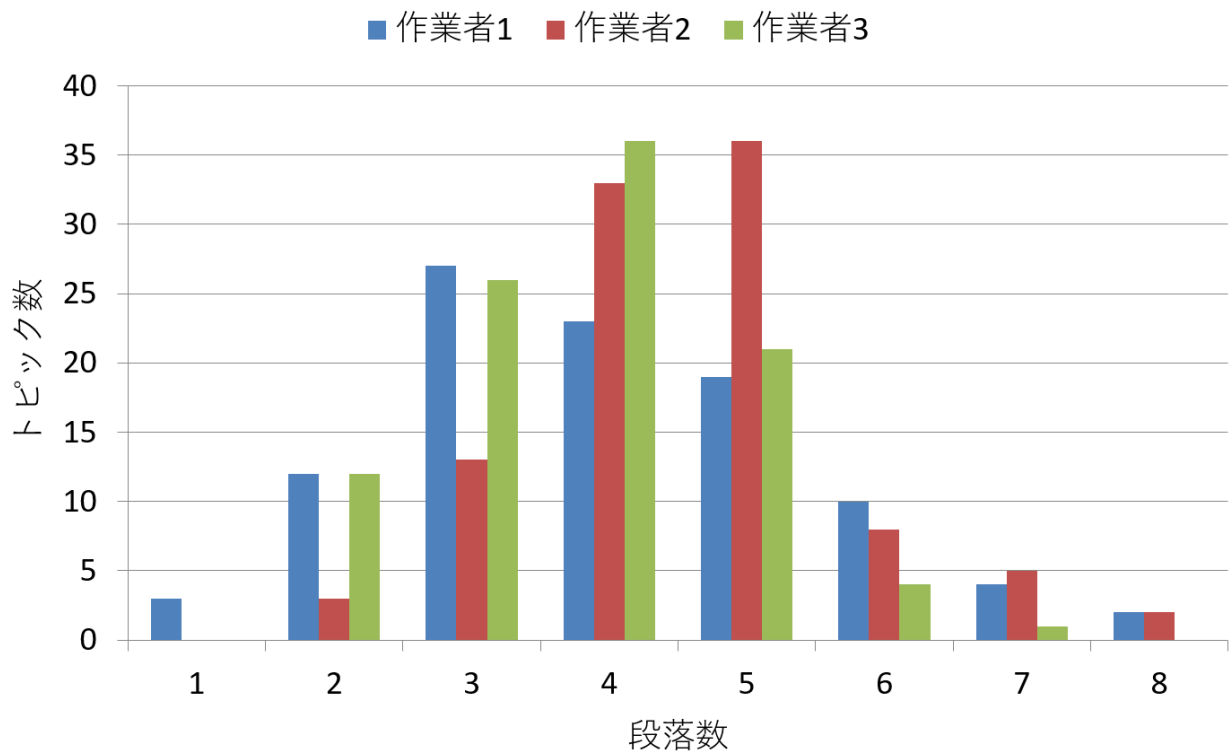


図 D.17: 重要段落として選ばれた段落の数のヒストグラム (要約 B)

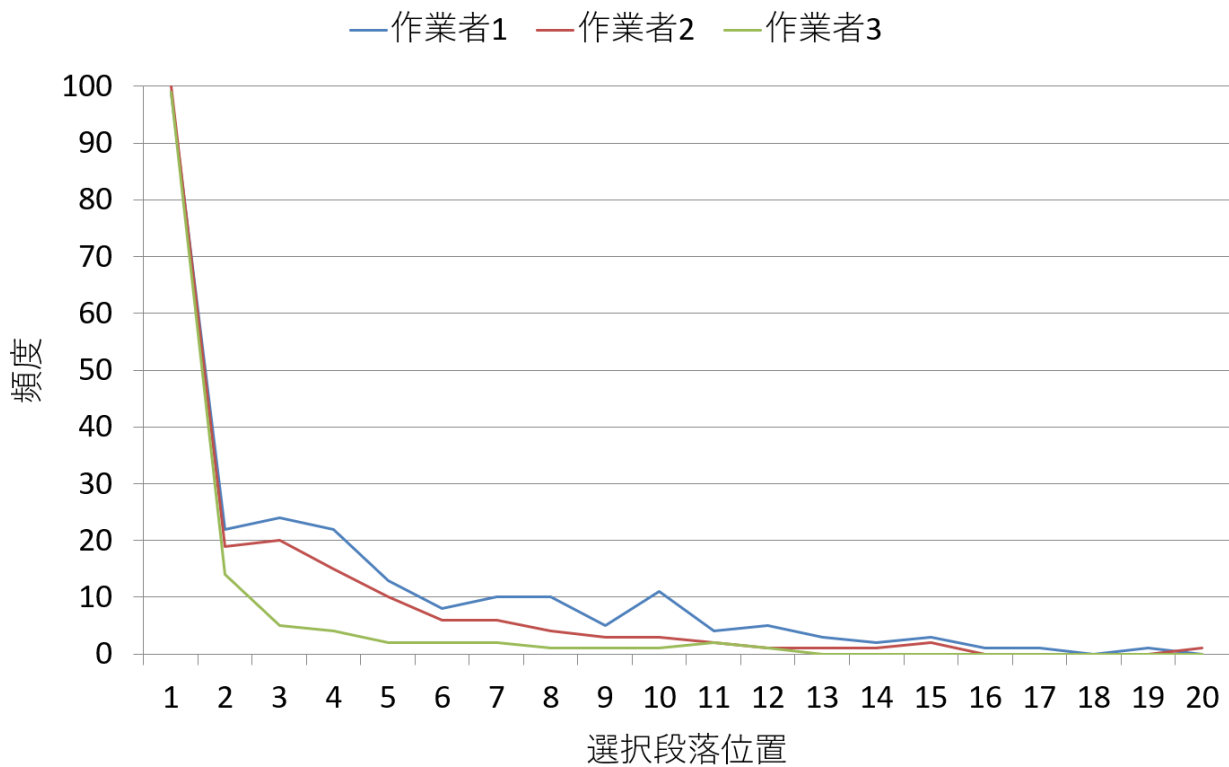


図 D.18: 重要段落位置のヒストグラム (要約 A)

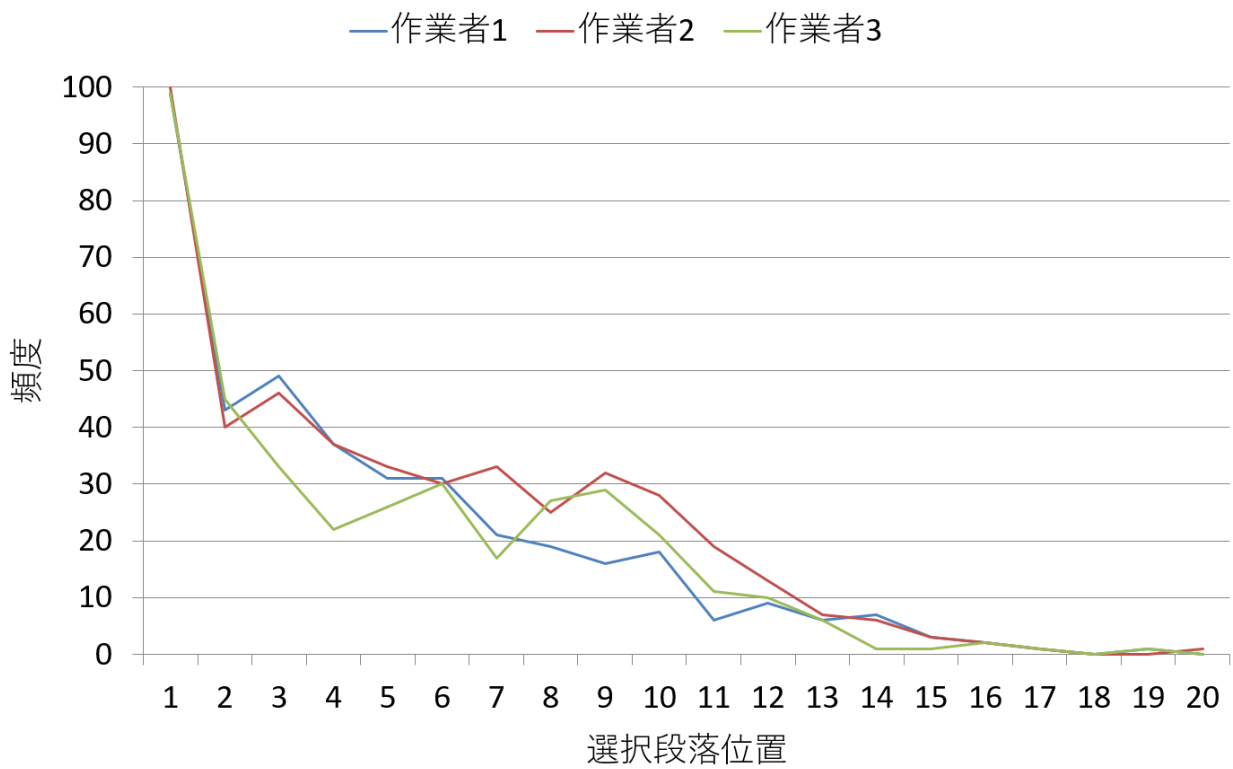


図 D.19: 重要段落位置のヒストグラム (要約 B)

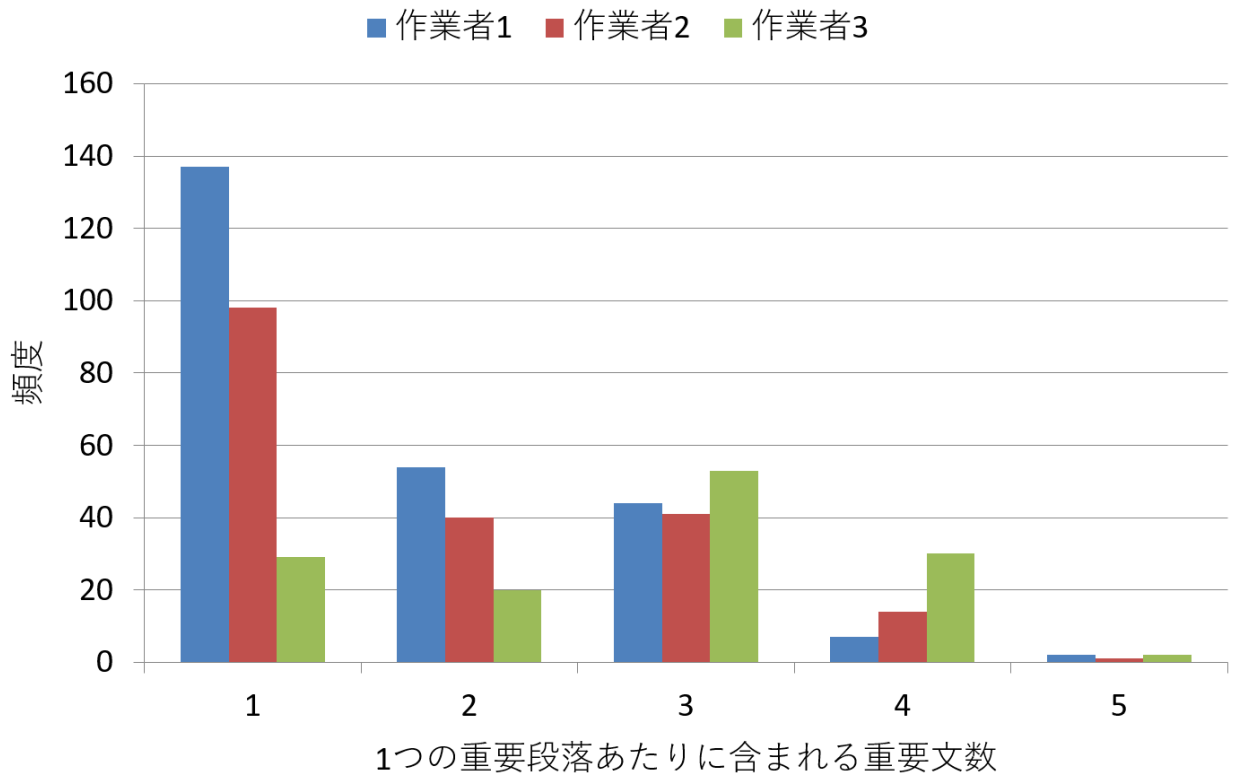


図 D.20: 重要段落が包含する重要文の数のヒストグラム (要約 A)

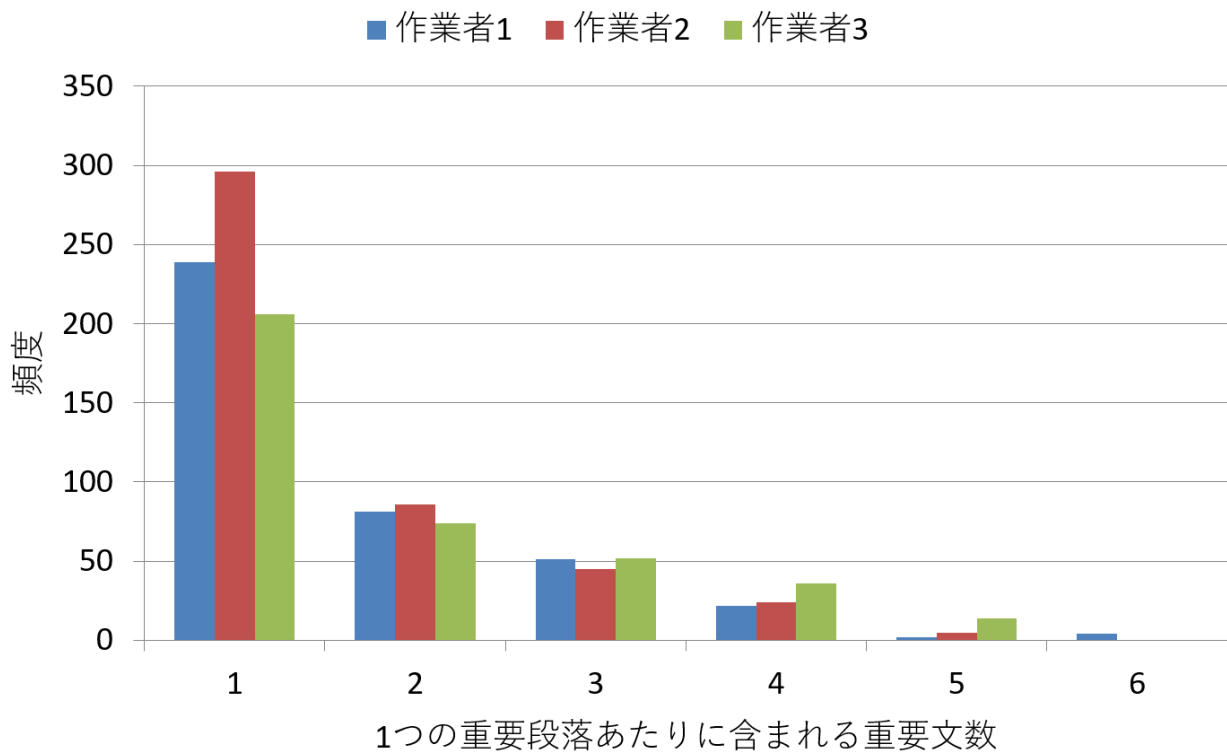


図 D.21: 重要段落が包含する重要文の数のヒストグラム (要約 B)

D.3 Yahoo コーパス vs. 日経コーパス

D.3.1 実験設定

重要文抽出モデルとして最大被覆モデル (MCKP) を使用した。Yahoo コーパスの重要文はタイプ a に相当する。評価は 10 分割交差検定で行う。コーパスの作業者 i_n の 10 トピックをテストとしたとき、作業者 j_n の残り 90 トピックで単語重要度推定器 (RandomForest) を学習し、ROUGE-1 で評価する。また、Yahoo コーパスの作業者 k_y の 100 トピックで単語重要度推定器を学習し、日経コーパスのテストセットに対して評価する。同様に Yahoo コーパスに対しても評価する。

D.3.2 実験結果

Yahoo コーパスで学習し，Yahoo コーパスで評価した結果を図 D.22 に示す．日経コーパス（要約 A）で学習し，Yahoo コーパスで評価した結果を図 D.23 に示す．これらを箱ひげ図で表したものを図 D.24 に示す．赤い棒は，訓練とテストで作業者が同じであるときの結果である．棒グラフの髭は 10 分割交差検定における ROUGE-1 値の最小値と最大値を表す．これらの結果から，日経コーパスで学習したモデルを Yahoo コーパスに適用しても性能が維持できていることが分かる．

日経コーパス（要約 A）で学習し，日経コーパス（要約 A）で評価した結果を図 D.25 に示す．Yahoo コーパスで学習し，日経コーパス（要約 A）で評価した結果を図 D.26 に示す．これらを箱ひげ図で表したものを図 D.27 に示す．日経コーパス（要約 B）で学習し，日経コーパス（要約 B）で評価した結果を図 D.28 に示す．Yahoo コーパスで学習し，日経コーパス（要約 B）で評価した結果を図 D.29 に示す．これらを箱ひげ図で表したものを図 D.30 に示す．これらの結果から，Yahoo コーパスで学習したモデルを日経コーパスに適用すると性能が落ちることが分かる．

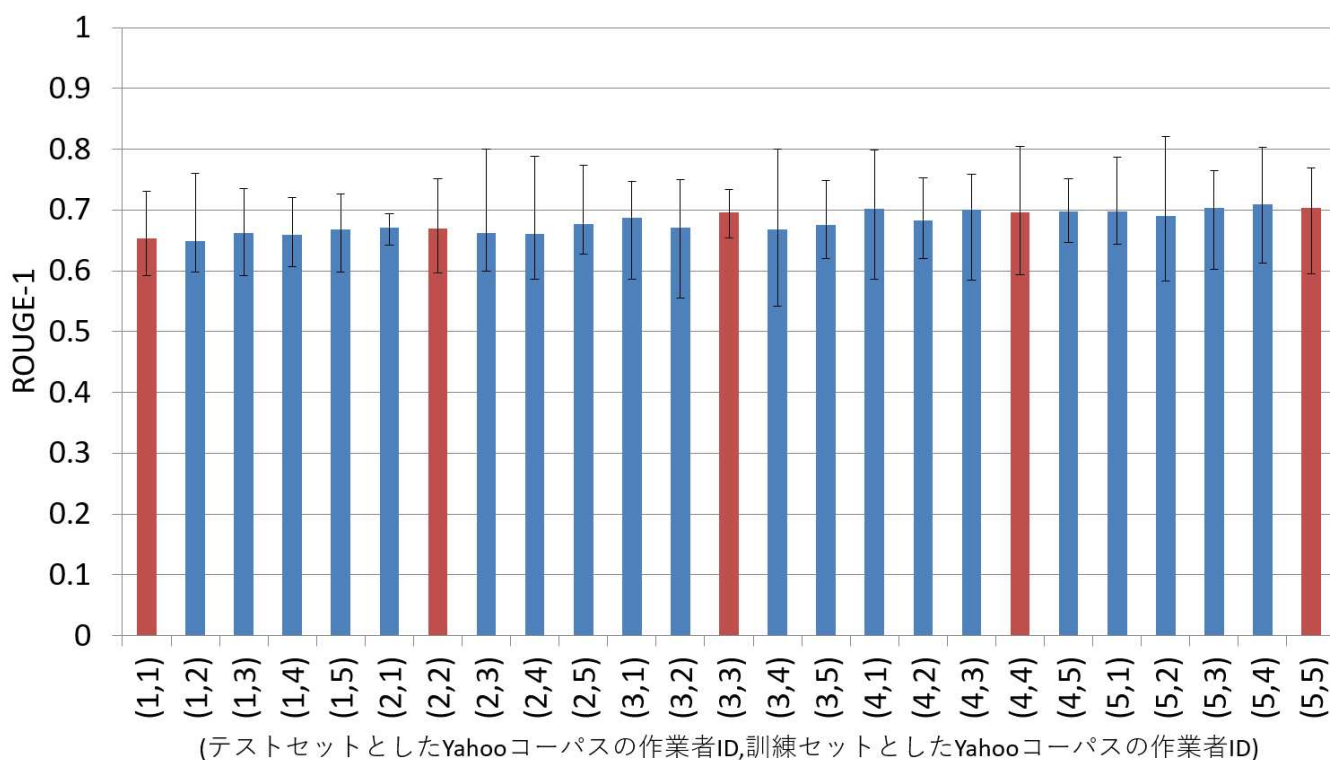


図 D.22: テスト：Yahoo コーパス，訓練：Yahoo コーパス

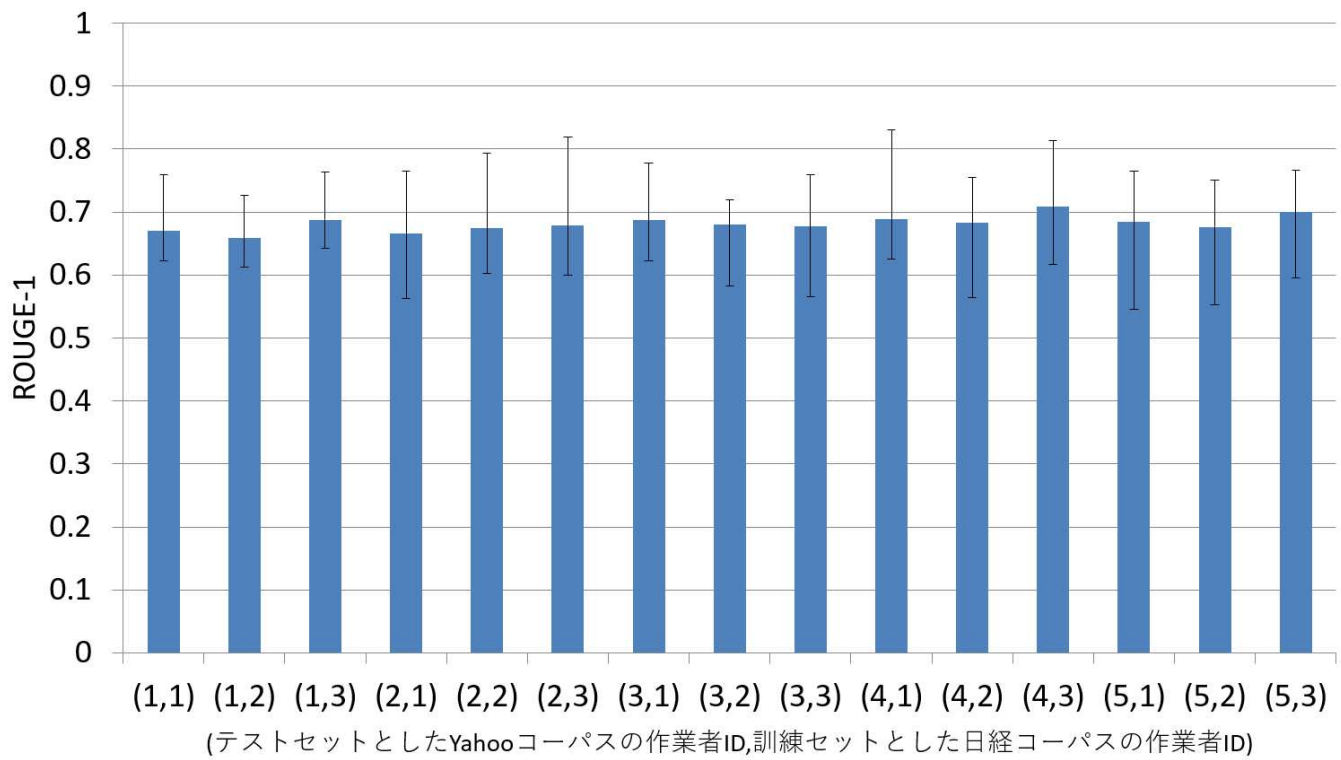


図 D.23: テスト : Yahoo コーパス, 訓練 : 日経コーパス (要約 A)

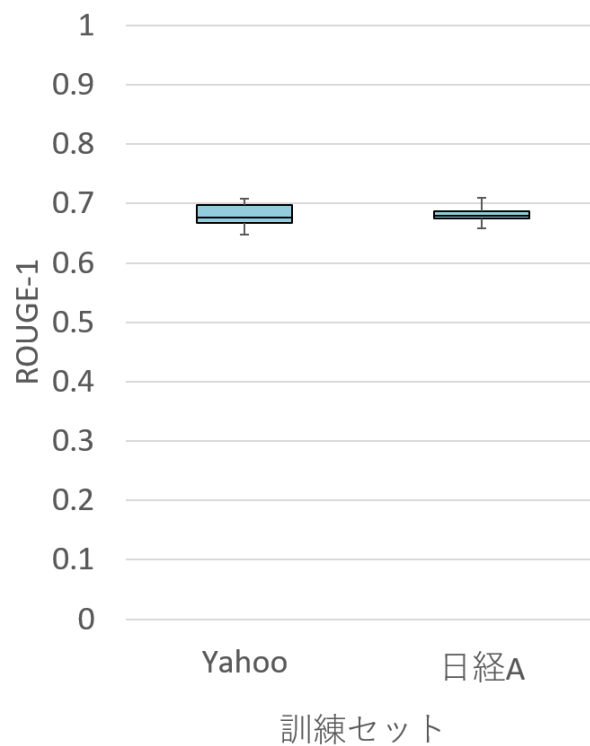


図 D.24: Yahoo コーパスをテストセットとしたときの結果比較

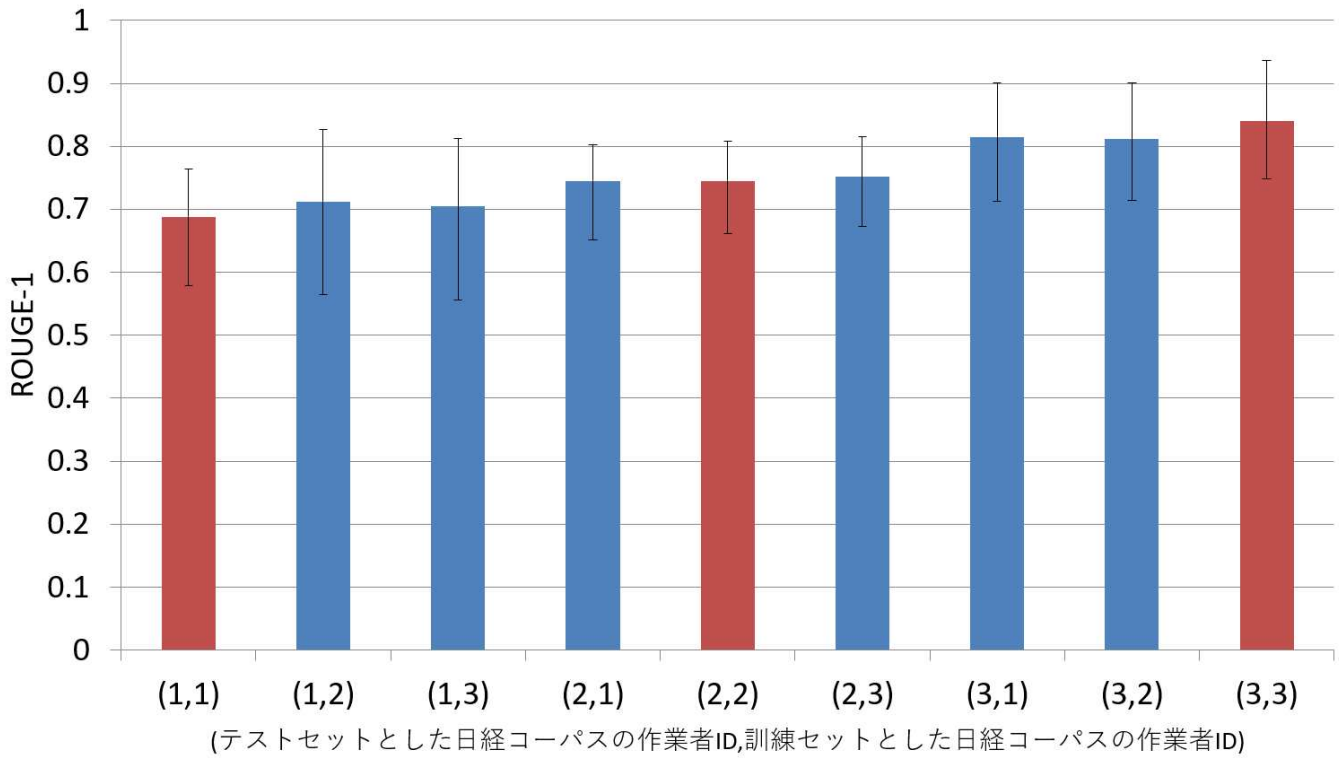


図 D.25: テスト : 日経コーパス (要約 A), 訓練 : 日経コーパス (要約 A)

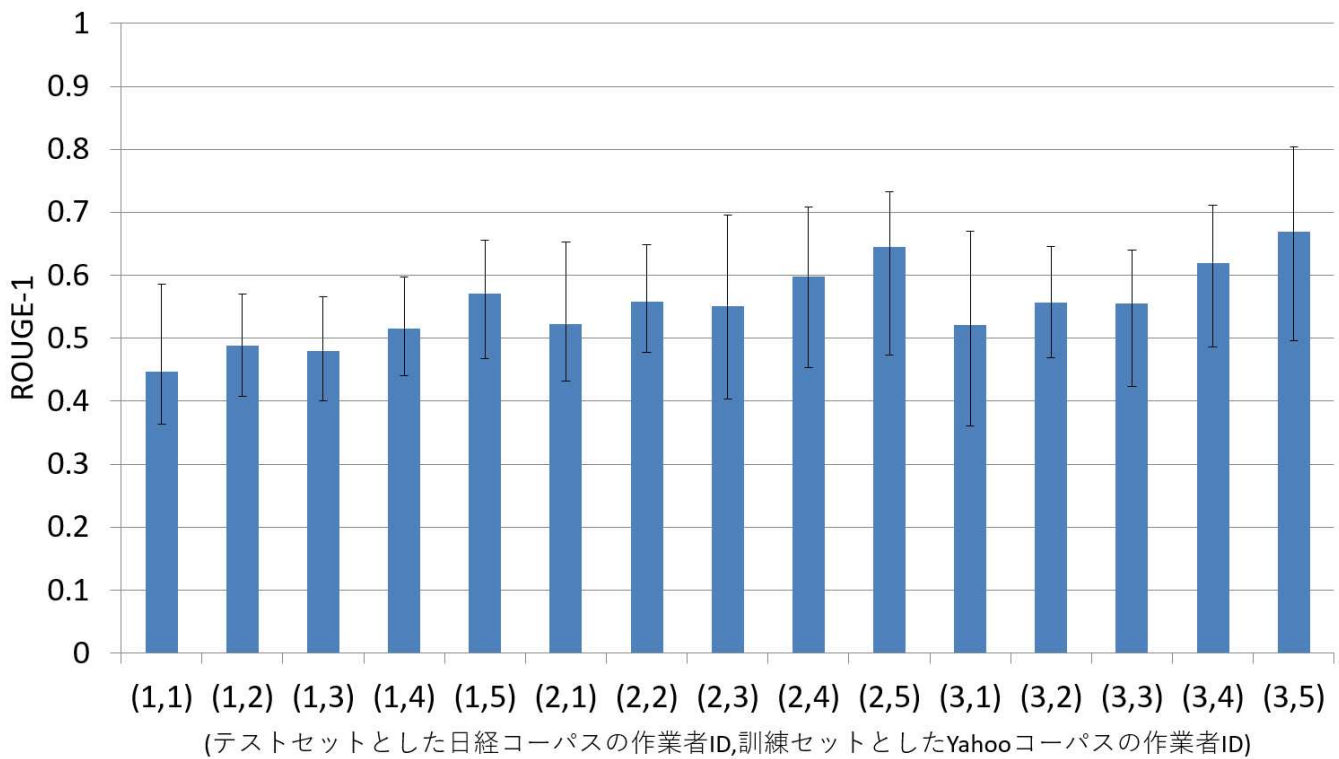


図 D.26: テスト : 日経コーパス (要約 A), 訓練 : Yahoo コーパス

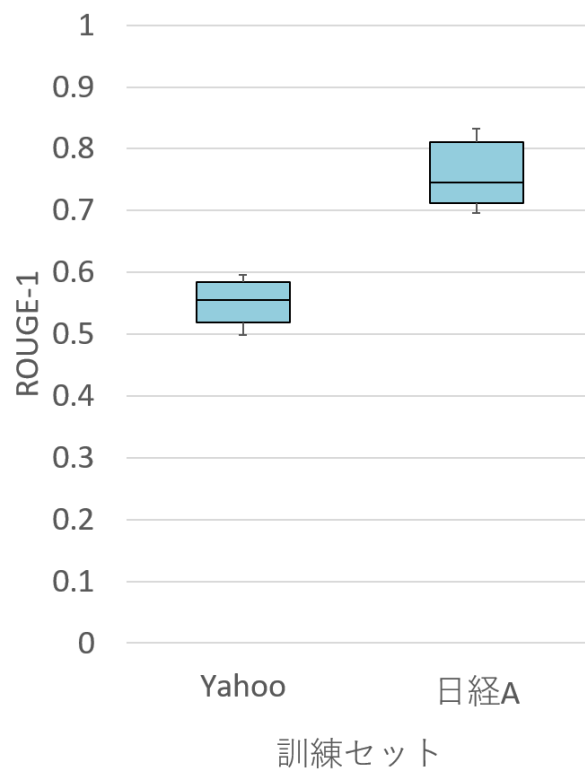


図 D.27: 日経コーパス（要約 A）をテストセットとしたときの結果比較

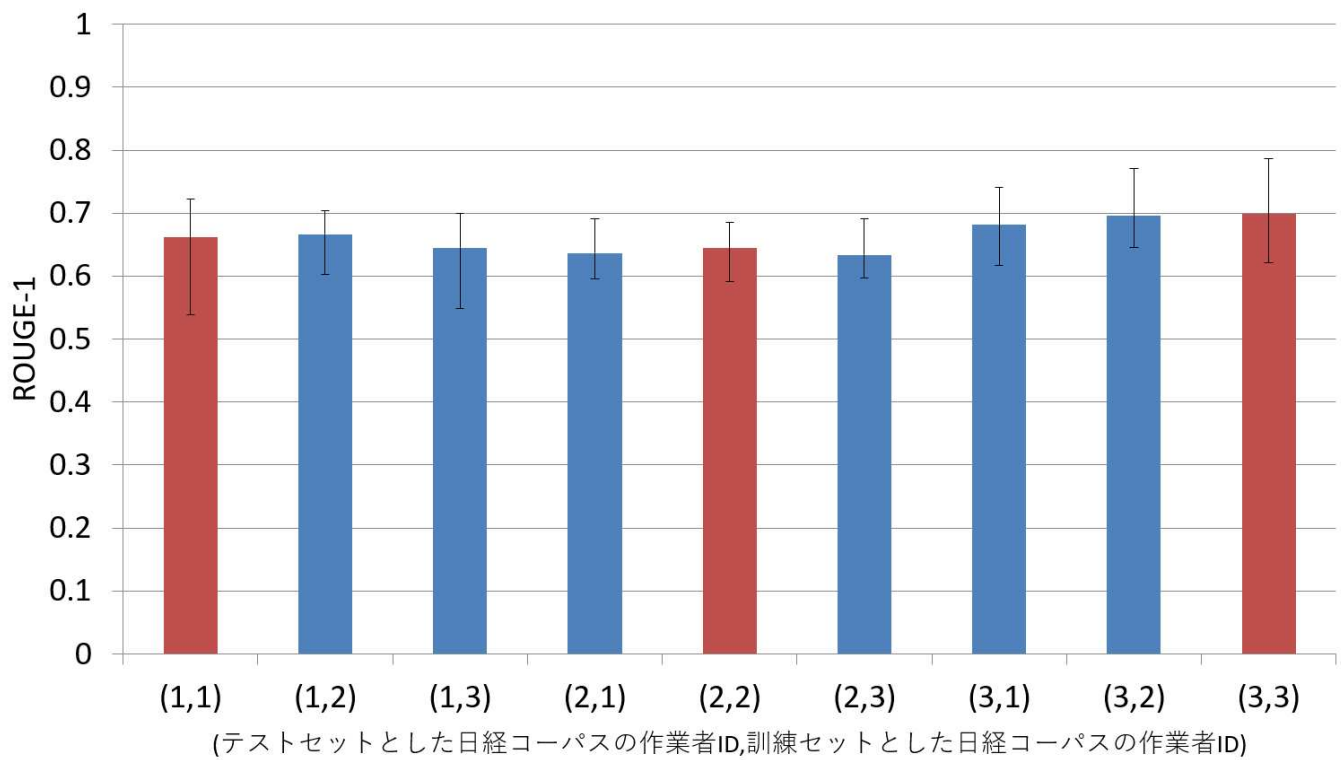


図 D.28: テスト：日経コーパス（要約 B），訓練：日経コーパス（要約 B）

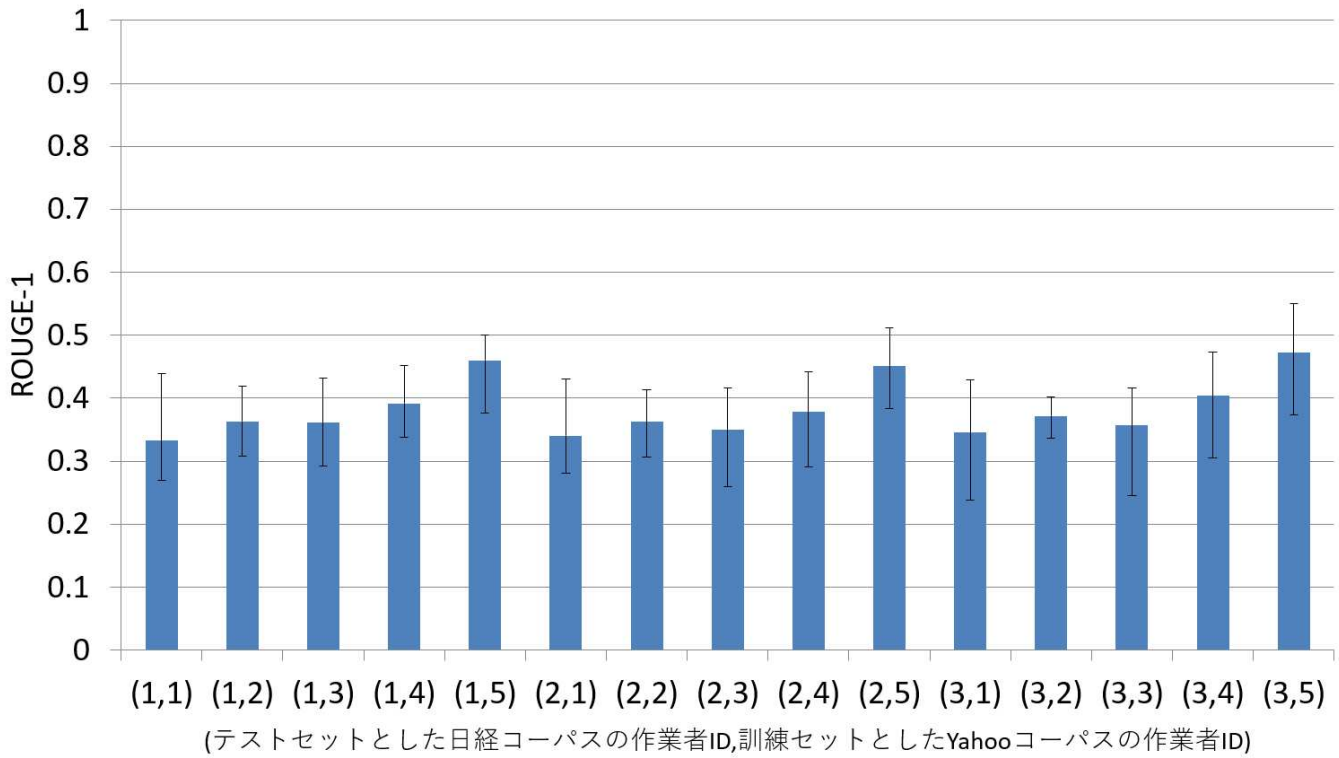


図 D.29: テスト : 日経コーパス (要約 B), 訓練 : Yahoo コーパス

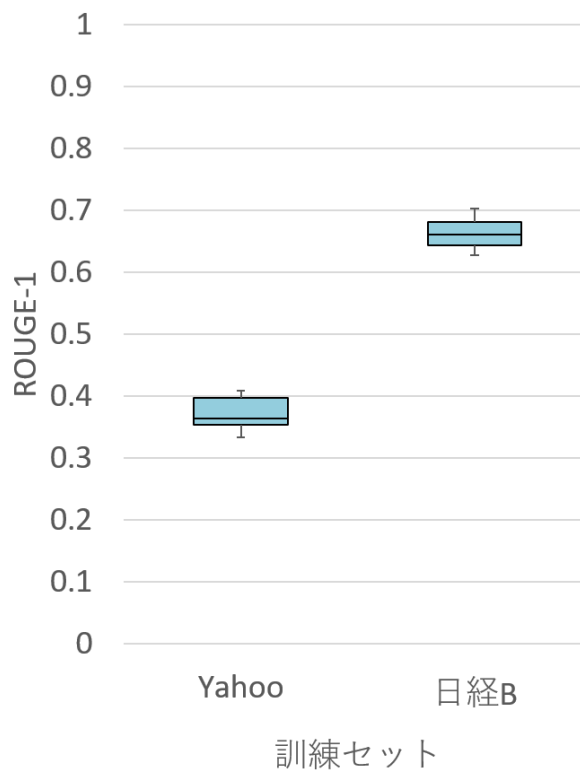


図 D.30: 日経コーパス (要約 B) をテストセットとしたときの結果比較

付 録 E 音声合成用音声データの分析

6.4.1 節で収録した音声合成用音声データについて，発話節の基本周波数（F0）や発話速度の変化に関する分析を行った [福岡 17a, Fukuoka 17b]．

E.1 分析条件

標本化周波数は 16kHz とした．F0 の抽出には STRAIGHT [Kawahara 06] を使用し，フレーム長 40ms，フレームシフト 1ms で抽出を行った．発話速度は 1 秒あたりのモーラ数と定めた．発話速度を計算するための音素アライメントには Julius¹ を用いた．

各発話節に対して，発話節全体の F0 平均，発話節全体の F0 ダイナミックレンジ，発話節冒頭の F0 平均，発話速度を算出した．

E.2 分析結果

発話節の位置と韻律の関係を図 E.1 に，発話節の役割と韻律の関係を図 E.2 に示す．

発話節全体の F0 平均の分布（図 E.1 (a)，図 E.2 (a)）を見ると，発話節の役割の影響が強く出ており，核となる発話節の F0 平均が高くなっている．これは発話節全体の F0 のダイナミックレンジ（図 E.1 (b)，図 E.2 (b)）でも同様である．

一方，発話節冒頭 3 モーラの F0 平均値を調べたところ（図 E.1 (c)，図 E.2 (c)），位置の影響が強くなっており，特に 1 番目の発話節冒頭では F0 が高い値を取っている．役割についても前置きの発話節冒頭の F0 平均が高くなっている．これは図 6.4 に示したように前置きの発話節の多くは 1 番目の位置に現れることに起因するものと考えられる．段落の最初の発話節冒頭で F0 が上がっている現象は，聞き手に注意を向けさせるための合図としての役割を果たしていると考えられる．

発話速度（図 E.1 (d)，図 E.2 (d)）については役割の影響が強く出ており，核の役割を持った発話節が他と比較してゆっくり発話されていることが見て取れる．位置との関係（図 E.1 (d)）を見ると，2 番目の発話節の発話速度が低下していることが分かる．これは図 6.4 に示したように核の発話節が 2 番目に多く現れることに起因するものと考えられる．

¹<http://julius.osdn.jp/>

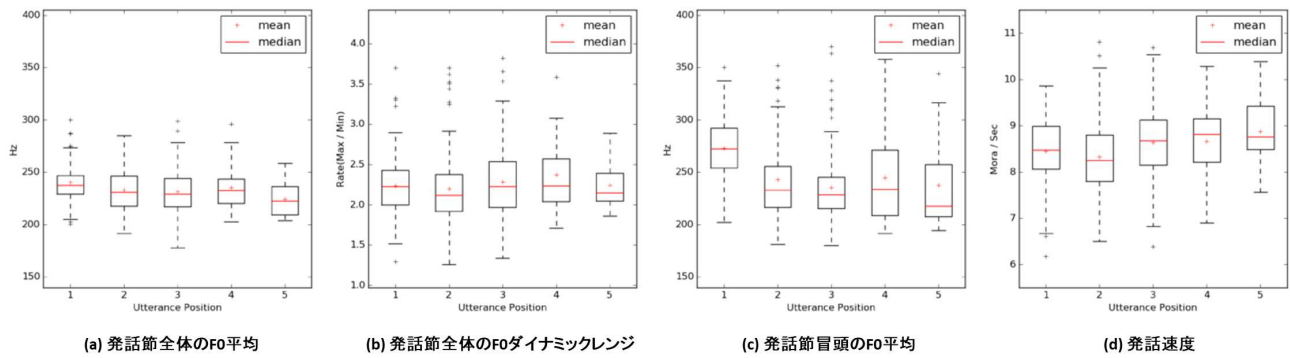


図 E.1: 発話節の位置による韻律の違い

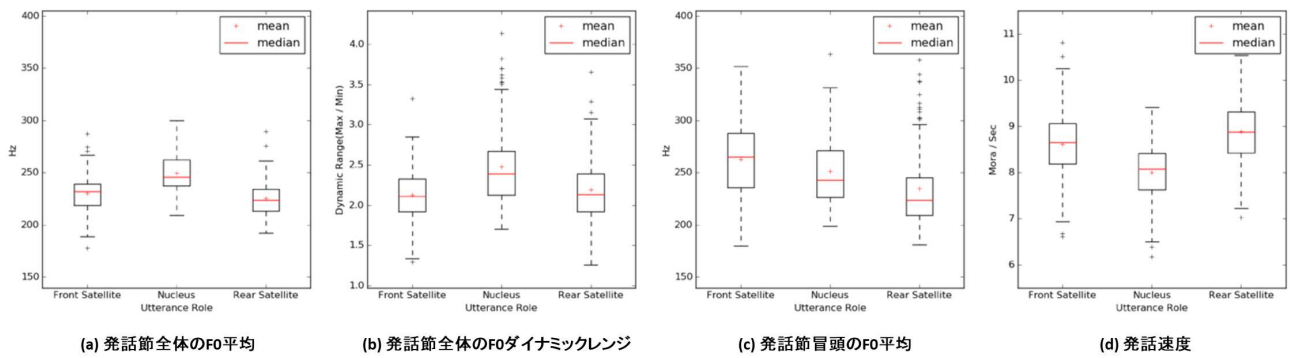


図 E.2: 発話節の役割による韻律の違い

付 録 F 要約・平易化の従来研究

F.1 自動要約の従来研究

Summarization とは、テキストの内容をより短いテキストで簡潔にまとめる処理である [奥村 05]。Summarization によってできたものを Summary という。日本語では、Summarization も Summary も要約と訳す。テキストから要約を自動で作成する処理をテキスト自動要約と呼ぶ。

テキスト自動要約の研究は、Luhn の研究まで遡る [Luhn 58]。Luhn らは、文に重要度を付与して重要文抽出を行うという手法で技術論文を対象に要約を行った。文書中で頻出する単語は重要だという考え方に基づいて重要語を決定した。ただし、“a”や“the”といったどの文書でも出現するような単語を取り除くため、頻度に上限と下限を定め、その区間に属する単語を重要語とした。重要語をベースとしたスコアは文を一定範囲 (bracket) に区切った単位で算出を行う。文からとれるいくつかの bracket のうち、値が最大のものを文のスコアとした。

文に重要度を付与して重要な文を抽出するという方法は、Luhn の研究以降も続けられた。重要文抽出で有効なテキストの特徴として、Paice は、単語の頻度、文の位置、タイトル、手がかり表現、文や単語間のつながり、談話構造を挙げた [Paice 90]。Edmundson の調査では、文の位置情報、手がかり表現、タイトル、単語の出現頻度の順に精度に寄与すると分析した [Edmundson 69]。特に、最初の 3 つを組み合わせると最も良い結果が得られると報告した。Zechner は、TF-IDF で単語を重み付けし、文中に現れる単語の重みの和を文の重要度として重要文抽出を行い、人間の被験者同士の比較結果と大差ないことを示した [Zechner 96]。Brandow らは、単語の出現頻度や見出し情報を利用した要約システム ANES を開発し、新聞や雑誌の記事を対象に、受容可能性 (acceptability) に関してリード法と比較したところ、リード法の方が優れていることを確認した [Brandow 95]。Lin らは、Ziff-Davis の新聞記事について、キーワードの数に基づいて文の重要度を計算し、どの位置の文が重要であるか調査した [Lin 97]。その結果、段落に関しては、2 番目と 3 番目の順で重要であり、段落内の文の位置に関しては先頭に近いほど重要であることを確認した。Watanabe は、単語の出現頻度や位置情報の他に、時制や文タイプ、前文との接続関係を用いて、これらの重みを重回帰分析で求め、文の重要度を計算した [Watanabe 96]。

1990年代後半頃から、文の重要度を計算するのに機械学習が使われ始めた。文の重要度の推定に、Kupiecらはナイーブベイズ分類器を [Kupiec 95], Linらは決定木を [Lin 99], HiraoらはSVMを [Hirao 02], Osborneは対数線形モデルを用いた [Osborne 02]。

テキスト中の多数の文と関連が強い文は重要であるという考え方に基づいて重要文抽出を行うグラフベースの手法がある。Skorokhod'koは、文をノード、文間の関係をリンクとしてグラフを作成し、多くの文と関連する文を重要文として抽出する手法を提案した [Skorokhod'ko 72]。Maniらは、テキスト中の単語をノードとし、単語間の隣接性、構文的関係、共参照関係、語彙的類似性などの関係をアークで表現したグラフでテキストを表現し、このグラフ中での活性値の伝搬により、高い活性値を得た単語や句、文を重要とみなす重要文抽出手法を提案した [Mani 97]。Saltonらは、段落をノードとして類似度が高い段落同士をリンクで結び、text relationship mapsを作成し、このグラフから重要段落を抽出する手法を提案した [Salton 96]。Mitraらは、百科事典を対象にSaltonら手法を評価し、2人の人間が抽出した重要段落の一致度と、システムと人が選んだ重要段落の一致度が同程度であることを確認した [Mitra 97]。Saltonらの手法は、多くのノードとリンクしているノードは重要であるとする次数中心性の考え方であった。一方、重要なノード（文や段落）からリンクが張られているノードは重要であるとする固有ベクトル中心性の考え方に基づいて複数文書要約を行う手法としてLexRank [Erkan 04]とTextRank [Mihalcea 04]がある。Barriosらは、TextRankの文間の類似度として様々な尺度を比較した。その結果、オリジナルの類似度よりも、BM25、TF-IDFベクトルのコサイン類似度、最長共通部分列一致の順で良いことを確認した [Barrios 15]。北島らは、LexRankの文間の類似度としてLDA (Latent Dirichlet Allocation) で推定したトピック分布の類似度を用いたTopicRankを提案した [北島 12]。

冗長性を考慮して文を抽出する手法として、MMR (Maximal Marginal Relevance) がある [Carbonell 98]。MMRは、ユーザーから与えられたクエリをもとに動的に要約を作成する。クエリと文の類似度から、文間の類似度を差し引いたスコアに基づいて重要文抽出を行う。MMRを複数文書要約用に拡張した手法としてMMR-MD (Maximal Marginal Relevance Multi-Document) がある [Goldstein 00]。MMRでは一つのテキストに含まれる文集合から重要文を抽出するのに対し、MMR-MDではクエリを検索システムに与え得られた文集合から重要文を抽出する。森らは、MMR-MDのクエリと文の類似度の計算に、質問応答エンジンに基づく語の重要度と情報利得比に基づく語の重要度を用いることで改善を試みた [Mori 04]。

Filatovaらは、重要文抽出を最大被覆問題として定式化した [Filatova 04]。最大被覆問題は、重要な概念（単語など）を制限要約長に収まる範囲でできるだけ多く含めるような文集合を選択する問題である。最大被覆問題では、冗長性に関して直接的な制約は設けていないが、制限

要約長内でできるだけ多くの内容を含めようとすることで、自然と冗長性が減ることになる。最大被覆問題は NP 困難な問題である [Khuller 99] デコーディング方法として、Filatova らは貪欲法を用いた [Filatova 04]。Yih らはスタック・デコーディング [Jelinek 69] を用いた [Yih 07]。Khuller らは乱択アルゴリズム [Hromkovic 03] を用いた [Khuller 06]。高村らは上記方法に加え、性能保証付貪欲アルゴリズム [Khuller 99] を取り入れ、比較を行った [Takamura 09]。西川らは、ラグランジュヒューリスティックに基づくデコーディング法を提案し、整数計画ソルバーよりも 100 倍以上高速であり、最大被覆問題の最適解と同程度の ROUGE を示すことを確認した [西川 13]。

Marcu は、テキスト構造を利用して重要文を抽出する手法を提案した [Marcu 98]。この手法では、まず、文章を EDU (Elementary Discourse Unit) と呼ばれる文や節などの談話の基本単位に分解し、修辞構造理論に基づいて、談話構造木 (RST-DT) を作成する。葉ノードにくる EDU を核 (nucleus:他の談話単位よりも重要と判断された談話単位) と衛星 (satellite:核に比べて重要でない談話単位) に分類し、核を親ノードに伝搬させていくことで、重要な情報と重要でない情報を区別する。談話構造木の上位階層ほど重要度が高い文として、要約率に達するまで文を抽出する。Hirao らは、談話構造木の刈り込みによる要約生成を行うために、RST-DT を EDU 間の依存関係が直接分かる構造として、依存構造に基づく談話構造木 (DEP-DT) へ変換する手続きを提案し、長さ制約のもとで談話構造木の根付き部分木を選択する問題として定式化した [Hirao 13]。しかしながら、Hirao らの手法は、EDU を文よりも小さい節単位で抽出するため、組み合わせ最適化問題として解いた場合、情報の断片化が生じてしまう。一方で、文を抽出単位とすると文レベルの文法性は保たれるが、不要な情報を含めてしまうという問題がある。菊池らは、文書を文間の談話構造木で表現し、文を単語間の係り受け木で表現し、その木構造を制約として部分木を選択する要約手法を提案した [Kikuchi 14, 菊池 15]。

最近では、機械翻訳分野での Sequence-to-Sequence (seq2seq) モデル [Sutskever 14] の成功を受け、要約分野でも Encoder-Decoder モデルを用いた要約手法が盛んに研究されている。Encoder-Decoder モデルを用いた要約手法には、抽出型 (Extractive) の手法と抽象型 (Abstractive) の手法がある。

抽出型のモデルでは、文や文書全体をエンコードして得られた埋め込み表現を用いて系列ラベリングの要領で文や単語を要約に含めるかどうかを識別する。Filippova は、seq2seq のモデルを用いた文圧縮手法を提案した [Filippova 15]。このモデルでは、Encoder に入力した文と同じものを Decoder の入力として与え、単語ごとに要約に含めるか含めないかのラベルを推定する。Cheng は、単語系列に対して CNN を適用し文の表現を得た後に attention+seq2seq モデルを用いて系列ラベリングにより文を要約に含めるかどうかを識別するモデル (NN-SE) と、

文に対する attention を計算した後に文の単語に対する attention を計算し Pointing Mechanism により単語を抽出するモデル (NN-WE) を提案した [Cheng 16]. Nallapati らは、文に対する双方向 LSTM から文書の表現 d を計算し、文を要約に含めるかどうかを判定する際に、毎回考慮するモデル (SummaRuNNer) を提案した [Nallapati 17]. 文を要約に含めるかどうかの確率は、該当文の双方向 LSTM の結合 h から計算される内容 (content), d と h の双線形写像で計算される顕著性 (salience), これまでの文系列の隠れ層を要約に含まれる確率で重み付けした和を非線形変換したものと h の双線形写像で計算された値を差し引いた新規性 (novelty), 文の先頭からの位置と長さで正規化した位置に関する特徴, およびバイアス項, これらの和に対してシグモイド関数を適用することで求める. Al-Sabahi は、文や文書の埋め込み表現を得るのに self-attention [Lin 17] を用いたモデルを提案した [Al-Sabahi 18]. 要約を含めるかどうかの確率は、Nallapati らの SummaRuNNer と同様に、内容、顕著性、新規性、位置に基づいて計算される.

抽象型のモデルでは、Encoder で文や文書をエンコードし、そのエンコード結果を用いて Decoder で言語モデルのように要約を生成する. そのため、抽出型では難しい、表現の言い換えや抽象化が行える. Encoder-Decoder モデルが登場する以前は、WordNet を用いて「John bought some vegetables, fruit, bread and milk → John bought some groceries」のように概念統合することで抽象化が行われていた [Hovy 97].

Encoder-Decoder モデルによる抽象型要約の研究は、Rush らの研究を源流とする. Rush らは、attention 機構を用いた抽象型要約モデルを提案した [Rush 15]. モデルの構造は Feed-Forward のニューラル言語モデルである. Chopra らは、このモデルを RNNLM に拡張したモデル (RAS; Recurrent Attentive Summarizer) を提案した [Chopra 16]. RAS の Decoder は、活性化関数がシグモイドの Elman 型の RNN または LSTM である. 一方、Encoder は、attention 機構を持つ convolutional network である.

機械翻訳や要約生成で用いられる Decoder では、事前に規定したサイズの辞書 (shortlist) を用いて単語を生成する. この shortlist を用いたアプローチには、低頻度語の問題と未知語の問題がある. 低頻度の単語は意味表現を学習するのが難しく、それらを扱おうとすると辞書サイズが肥大化し、学習のパフォーマンスが低下する. そこで、低頻度語を UNK と呼ばれる特別な記号に置き換える処理がしばしば行われるが、これにより、様々な重要な情報が欠落してしまう. これらの問題を軽減する方法として、入力情報を陽に参照する方法が考えられる [Vinyals 15a]. 入力情報を活用することで shortlist に含まれない単語を生成することができる. 実際、要約問題では、与えられた情報の一部をそのまま参照して新しいテキストを生成することが多い. Gulcehre らは、コピーメカニズムを導入することで低頻度語および未知語に頑

健な Encoder-Decoder モデルを提案した [Gulcehre 16]。この手法では、従来の shortlist に基づく softmax (Shortlist Softmax) の他に、入力文の各単語の位置に関する softmax (Location Softmax) を利用する。Shortlist Softmax は、shortlist に含まれる単語の Decoder での出力確率である。一方、Location Softmax は、attention 計算時に使用する Encoder の各隠れ層に対する荷重である。この値を確認することで、入力のどの単語が参照されやすいかが分かる。そして、それらのうち、どちらの Softmax を重視するかを Switching network が決める。Switching network は、attention の計算で求めた文脈ベクトルと Decoder の前時刻の隠れ層を入力とする多層パーセプトロンでモデル化され、その出力を sigmoid gate に与えることで、重みが求まる。そして、このスイッチング確率で重みづけた各 Softmax の出力ベクトルの結合を Pointer Softmax と呼ぶ。この仕組みにより、shortlist に含まれない未知語であっても生成することが可能となる。ただし、shortlist にも入力にも含まれない単語に関しては従来通り UNK を生成する。このようなコピーメカニズムを導入した Encoder-Decoder モデルが、現在の抽象型要約のモデルとして主流になりつつある [Nallapati 16, See 17, Paulus 18, Celikyilmaz 18]。

機械翻訳タスクは「損失なし生成」(loss-less generation) 問題であり、要約は「損失あり圧縮」(lossy compression) 問題である。Nallapati らは、attention 付き Encoder-Decoder モデルにおいて、翻訳タスクで検討されている様々な知見が要約タスクでも有効か、以下の4つの観点について検討を行った [Nallapati 16]。(1) Large vocabulary trick [Jean 14]: decoder の語彙をミニバッチ内の語彙と高頻度語の語彙を規定の数になるまで取ってくる手法。(2) Feature-rich Encoder: Encoder の入力として単語だけでなく、品詞や固有名詞、頻度や IDF などの補助的な情報を加える。(3) Pointing Mechanism (Copy Mechanism) [Gulcehre 16]: 語彙リストから生成するか入力に含まれる単語を使用するか選択できる枠組みを導入する。Switch が on のときは語彙リストから出力し、off のときはソースから単語を取ってくる。switch が off になるのは、未知語のときである。(4) 階層的な attention: 複数文入力で与え、単語に対する attention だけでなく、文に対する attention も行う手法を導入する。これら4つの仕組みを加えることで、デフォルトの attention 付き Encoder-Decoder モデル(1文入力、素性は単語のみ)よりも高い性能を示すことを確認した。

Nallapati らの Pointing Mechanism モデル [Nallapati 16] では、語彙リストから生成するか入力に含まれる単語を使用するかをスイッチで切り替え、どちらかの分布を使用していた。See らは、Pointer 分布と Generator 分布の足しあわせで最終的な分布を表現する Pointer-Generator モデルを提案した [See 17]。未知語の場合は Generator 分布はゼロとなり、Pointer 分布のみから出力を決定する。前時刻までの attention の和を coverage vector と定義し、今まで本文のどこを見ていたかを明示的に与えることで、単語の繰り返し生成を防ぐ手法として、Coverage

Mechanism を提案した．従来の attention 計算は対象の Encoder の隠れ層と現在の Decoder の隠れ層を用いて行われるが，このモデルではさらに coverage vector を加えた式から計算される．

Encoder-Decoder に基づく要約モデルにおける最尤法の学習には，テスト時 (Free running) の生成系列が学習時 (Teacher forcing) から外れだすとエラーが蓄積し続ける問題 (Exposure Bias) と，生成結果が評価指標である ROUGE に最適化されていない問題がある．これらの問題の対策として，Paulus らは，Self-Critical Sequence Training [Rennie 17] という強化学習の手法を用いた [Paulus 18]．Self-Critical Sequence Training では，ベースラインとして現時点のパラメータで生成した最良の系列に対する報酬を使用しており，ベースラインとして平均報酬を用いる手法 [Ranzato 16] よりも勾配の分散が抑えられ，学習が安定するとされる．しかしながら，この強化学習に基づく損失関数は，報酬関数である ROUGE に対して最適化され，要約に含めるべき内容は充実するが，可読性を保証するものではない．そこで，Paulus らは，可読性の改善も行えるように最尤法の損失関数と組み合わせた損失関数を定義した．通常の attention 付き Encoder-Decoder モデルでは，Encoder 側の隠れ層に対して attention を計算するが，Paulus らのモデルでは，Decoder でこれまでに生成された単語に対しても attention を計算する (Intra-Decoder Attention)．また，Encoder 側に対する attention として，Temporal Attention [Sankaran 16] を用いた．通常の attention 計算では，入力方向に対して softmax をとり，現在のデコーディングステップでどれだけその単語に注目するかを計算する．一方，Temporal Attention では，過去のデコーディングにおいてこれまでにその単語どれだけ注目してきたかということも考慮する．

要約タスクでは入力文を全体を読んだうえで重要な情報のみをまとめ上げる必要があり，入力文が長くなりがちとなる．Celikyilmaz は，パラグラフごとにエージェントを割り当て、各エージェントを相互に連携させながらエンコーディングを行う手法を生成型要約モデルに導入した [Celikyilmaz 18]．ここで，各エージェントの Encoder は多層の双方向 LSTM で構成される．下層の情報を上層に渡す際，他のエージェントが下層で行ったエンコード結果の平均も一緒に上層の渡す．

ユーザーが望む要約の長さは，把握したい情報の粒度や要約を読むデバイスの大きさなど，利用状況に大きく依存する．Kikuchi らは，Encoder-Decoder モデルに出力長を制御する仕組みを導入することで，ユーザーが希望する長さの要約を作ることができる [Kikuchi 16]．*LexEmb* と呼ばれる手法では，LSTM への入力に入力単語の埋め込みだけでなく，残りの要約長の埋め込みベクトルも与える．そして，残りの長さが 0 以下になったところで生成を止める．*LenInit* と呼ばれる手法では，出力長に応じて LSTM のメモリセルを初期化する．通常の Encoder-Decoder モデルでは，Decoder の初期化時に Encoder の最終状態の隠れ層とメモリセルの値が与えられる．しかしながら，本手法では，隠れ層はそのまま引き継がれるものの，メ

モリセルの値は、長さに関する基底ベクトルと要約長（スカラー）の積が与えられる。そのため、本手法では、要約長の入力 は Decoder の初期化時に一回だけ行う。そして、後どれくらいの要約を生成できそうかという情報は LSTM の内部状態として管理する。ビームサーチの上位数件を観察してみると、指定した要約長 ± 数バイト程度の誤差に収まっていた。

マルチタスク学習により要約の性能改善を図る研究が行われている。マルチタスク学習は、複数のタスク間で共通する特徴がある場合、それらを同時に学習することで、単一のタスクに過学習されず、汎化性能の向上が期待できる。Isonuma らは、文書分類とのマルチタスク学習により重要文抽出の性能が向上することを示した [Isonuma 17]。Klerke らは、視線情報を活用することで文圧縮モデルの性能が向上することを示した [Klerke 16]。視線情報として eye-tracking コーパスを用いた。eye-tracking コーパスは、人間が文章を読むときの目の動きを収録したコーパスである。評価指標として、First pass duration と Regression duration を用いている。First pass duration は、文書を読む際、単語を初めて見たときに理解に費やす時間を表しており、文章における単語の難しさを表していると考えられる。一方、Regression duration は、再度読み返した際の各単語に視線をとどめた時間を表しており、人間が文章の意味を理解するうえで、各単語がどれくらい重要かという単語の重要度を表していると考えられる。これらの値は、データセット中では、平均からのばらつきをもとに 0 から 5 に離散化されている。これらのデータと文圧縮のデータをマルチタスク学習する手法を提案した。

自動要約の研究は、新聞記事や論文に限らず、様々なドメインのテキストに対して行われている。堀らは、講演音声を対象とした自動要約手法を提案した [堀 01]。各発話文から重要な単語を抽出し、それらをつなぎ合わせることで要約文を作成する。要約らしさの計算に、単語の頻度や係り受け構造に基づく単語の遷移確率の他、音声認識で得られる音響的・言語的信頼度を使用している。講演音声には、フィラーや言い直しなどの冗長表現が多く存在するため、幅田らは、冗長表現を不要箇所として削除することで、情報を欠落させずに要約を行う手法を提案している [幅田 01]。対話の書き起こしの文を MMR でランキングし、制限要約長までテキストの順序で文を出力する手法を提案している [Zechner 01]。Reithinger らは、日程調整やホテル予約などの交渉対話を対象とした要約手法を提案した [Reithinger 00]。話し手の全員が合意した内容に関心があるという交渉対話における前提を利用して、提案された内容のうち受理されたものを特定し、それを含めるように要約を作成している。山村らは、複数人対話の履歴を対象に、発話速度や発話タイミングなどの非言語情報を考慮した重要文抽出手法を提案した [山村 15]。星野らは、対話履歴に対して句構造解析を行い、生成された木構造に対してクラスタリングを行うことで要約を作成する手法を提案した [星野 18]。

F.2 平易化の従来研究

テキスト平易化はテキストの複雑さを減らし、テキストの分かりやすさと読みやすさを向上させるための技術であり、構文や語彙を修正することで実現される。Shardlow の調査によれば、Google Scholar にて “Text Simplification”, “Lexical Simplification”, “Syntactic Simplification” を検索すると 1994 年から 2014 年の 20 年間で平易化に関する論文の発行数が年々増加しているとされる [Shardlow 14]。特に、2010 年以降の研究では資源と手法が豊富になったことにより、この分野は急成長している。主に Simple Wikipedia に注目した資源が増加している [Yatskar 10, Coster 11, Naples 10]。手法に関しては、文脈を考慮し、語義曖昧性を解消しつつ、簡単な語彙に書き換える研究が多い [Biran 11, Belder 10]。

古くから、第二言語学習者のために、手動で平易化する研究が行われてきた [Blum 78]。一方、自動で平易化する取り組みとしては、文法チェッカーが最初であった [Hoard 92]。これは、Boeing 社で商用航空機マニュアルの記者のために開発されたものである。英語を母国語としない人でも文書の内容を容易に理解できるようにするための規格として ASD-STE100 (ASD Simplified Technical English Specification)¹がある。記者はこれに従ってマニュアルを作ることになる。そのマニュアル作成支援のために開発されたのが、この文法チェッカーである。同じ時期に、制限言語のために自動で平易化する研究も行われた [Adriaens 95]。これ以降、構文の平易化 [Chandrasekar 97] や語彙の平易化 [Devlin 98] へ拡張されていった。

語彙の平易化は、難しい語や句を検出して、平易な表現に置き換える手法である。まず、難しい語を検出し、次に、言い換え候補を生成する。そして、語義曖昧性解消を行った後、難易度に基づいてランキングし、一番簡単な語で置き換えることによって実現される。同義語を頻度でランキングする手法は、Devlin らのものが最初である [Devlin 98]。彼らは、WordNet から得た同義語のうち、最も頻度が高い語に置き換える手法を提案した。頻度には、Brown Corpus から得られた英語の代表的な頻度である Kucera-Francis frequency を使用している。現在でも同義語を頻度でランキングする手法はよく用いられるが、Brown Corpus よりも大規模なコーパスが使用されている。また、近年でも高頻度語を用いることで、読みやすさが改善することが確認されている [Rello 13]。さらに、普通の Wikipedia よりも Simple Wikipedia (平易なテキスト) から計算した頻度の方が有効に働くことが確認されている [Kauchak 13]。

単純な語彙換言システムの大きな問題として語義曖昧性解消の問題がある。この問題に対する一つのアプローチとして潜在言語モデル (LWLM; Latent Words Language Model) [Deschacht 09] を用いるのが有効ではないかと Shardlow は主張している [Shardlow 14]。潜在言語モデルは、入力語と意味的に関連のある単語集合を生成するような言語モデルである [Deschacht 09]。

¹<http://www.asd-ste100.org/>

言い換えステップでこのモデルを使うことで、周りの文脈を考慮して言い換え可能な語がある程度取れるのではないかと推察している。Glavas らは、Google n-gram から GloVe [Pennington 14] で獲得した分散表現を利用し、ベクトル空間上でのコサイン類似度に基づいて、換言対象語候補を獲得した [Glavas 15]。この自動獲得した換言対象語候補を用いて語彙の平易化を行った結果、従来の平易化コーパスを使った手法と大差ない性能であることを示した。しかしながら、分布類似度に基づいた換言を行っているため、cool を warm に置き換えてしまうような誤りが見られた。

平易化の手法として、WordNet の同義語に置き換える研究がよく行われている。Thomas らは、上位下位関係からグラフを作って、語彙平易化を行う手法を提案している [Thomas 12]。この手法では、文書に存在する内容語の synset を推定して、それを上位語に置き換え、文書全体の語彙数を減らすことで、語彙の平易化を行う。

文脈ベクトルを用いた語義曖昧性解消の手法もある [Biran 11, Bott 12]。Biran らの手法では、まず、全ての単語についてコーパスから周辺を集めておいて、共起語などから文脈ベクトルを作る。次に、入力文のベクトルとその単語の文脈ベクトルが似ていれば、その単語はその文脈で使える判断する。単語から句、句から単語、句から句への変換を行うためには、句単位の換言知識が必要である。そこで、Yatskar らは、Wikipedia の編集履歴から、単語から句、句から単語への置き換え情報を抽出する方法を提案している [Yatskar 10]。

テキスト平易化を同一言語内の翻訳タスクと考え（入力：難解な語彙と文法を持つ文，出力：平易な語彙と文法を持つ文），統計的機械翻訳の手法を用いた平易化が，2010 年頃から盛んに行われている。ここで，統計的機械翻訳のツールとしてほとんどの研究で Moses [Koehn 07]² を使用しており，対訳コーパスは，既存のアラインメント技術を用いて，English Wikipedia と Simple English Wikipedia から作成している [Coster 11, Zhu 10, Wubben 12]。Xu らは，人間の評価と相関がある平易化の自動評価尺度を提案し，統計的機械翻訳モデルを平易化尺度に対して最適化する方法を提案した [Xu 16]。ここで，目的関数として BLEU, FKBLEU, SARI を用いて統計的機械翻訳モデルを最適化し，人手による評価結果を比較したところ，平易さに関しては SARI が最も人手の評価結果と高い相関を示した。Stajner らは，フレーズベースの統計的機械翻訳手法を用いたテキスト平易化において，学習データの量や質がどのように影響するか分析を行った [Stajner 15]。量に関しては，学習データの量が多いほど BLEU 値が高くなる傾向が見られたが，その差は僅かであった。量の観点では，まず，データセットとしてノーマルな文とシンプルな文で BLEU 値を計算し，BLEU 値の区間に基づいてデータを分類した。その結果，BLEU 値の高い文対で学習したモデルの方が高い性能を示した。このことから，テキス

²<http://www.statmt.org/moses/>

ト平易化のためのフレーズベースの統計的機械翻訳では、量よりも質が重要だと考えられる。

英語以外の言語では、Simple English Wikipedia が使用できないため、専門家が新聞記事を手で平易化したコーパスを用いた研究もある [Specia 10]。また、単に Moses を使うのではなく、テキスト平易化に特化した翻訳手法も提案されている [Coster 11]。Coster らは、English Wikipedia と Simple English Wikipedia を比較し、テキストの平易化では句の削除が頻繁に起こることを発見した。そして、統計的機械翻訳の手法の中に句の削除という機能を追加し、BLEU による評価を行ったところ、デフォルトの Moses よりも高い BLEU 値が得られることを確認した。

語彙平易化における readability の自動評価では、Belder らによって作成された語彙平易化評価セット [Belder 12] を用いることで、語彙平易化手法の適合率と再現率と F 値を自動で評価することができる。一方で、機械翻訳手法の評価では、BLEU などの一般的な機械翻訳の評価尺度を用いることができる。しかしながら、これらは文単位での評価にすぎないので、テキスト全体での readability を評価する方法を検討する必要がある。

テキスト平易化の研究は、機械翻訳やテキスト生成、要約や言い換え生成などの研究と密接な関係がある。例えば、要約のコーパスアラインメント技術を平易化技術に用いた研究 [Barzilay 03, Bott 11] や、統計翻訳の技術を用いた研究がある [Zhu 10, Specia 10]。また、統計翻訳の評価尺度を平易化の評価尺度として使っている研究もある [Doddington 02]。要約は、文の長さを短くしたり、内容を減らすことを目的としている。一方、テキスト平易化でも一般的には平易なテキストは短くなる傾向にあるが、説明生成を伴う場合、長い文が生成されることもある。また、要約と異なり、基本的に内容は全て保持される。逆に、要約システムの中でも平易化の技術が利用されている研究もある [Blake 07, Silveira 12, Siddharthan 04]。これらの研究では、最終的な要約の読みやすさを向上させる目的で平易化技術を用いている。このように平易化技術は要約などの技術と密接に関係しているため、関連分野の技術が向上すると平易化技術も向上し、平易化技術の向上すれば関連分野も向上する。「平易さ」(simplicity) の定義は難しく、典型的には、平易さの評価では、文の長さ [Kincaid 75] や音節の数 [McLaughlin 69] などの表層情報が用いられる。一般に平易な文は短いので文の長さなどを用いることでそこそこの推定はできるが、常に正しいわけではない(説明生成を伴う場合などは長文の方が理解しやすく平易な場合がある)。

テキスト平易化技術が、製品化されたものは少ない。製品として使われた平易化技術として [Hoard 92] や [Scarton 10] がある。テキスト平易化技術が製品化されにくい主な理由は精度の低さが挙げられる。例えば、情報検索であれば、検索精度が低くてもユーザーは探している情報の一部を見つけることができる。また、精度の低い検索システムでも、それがなければユー

ザーは自力で情報を見つけるのはできない。一方で、平易化システムが平易でないテキストを出力した場合、ユーザーはそのテキストを理解することができなくなってしまう。そのため、誤りを含む文章よりかは、複雑であっても原文の方が理解しやすいということから、精度の低い平易化システムは製品化されにくい状況にある。

実際に作成された日本語の平易化システム・難易度判定システムとして、日本語文章難易度判定システム『jReadability』³、日本語テキストの難易度判定ツール『帯』[佐藤 08]⁴、日本語読解学習システム『リーディング・チュウ太』[川村 12]⁵、やさしい日本語チェッカー [山本 17]⁶、語彙平易化システム [梶原 17]⁷がある。

³<http://jreadability.net/>

⁴<http://kotoba.nuee.nagoya-u.ac.jp/sc/obi3/>

⁵<http://language.tiu.ac.jp/>

⁶<http://box.jnlp.org/easy-japanese/checker>

⁷http://moguranosenshi.sakura.ne.jp/lexical_simplification/

付 録 G 言語資源・言語処理ツール

本研究で使用した，または今後本研究を進めるうえで役立ちそうな，言語資源・言語処理ツールについて紹介する．

G.1 言語資源

G.1.1 ALAGIN の言語資源

高度言語情報融合フォーラム（Advanced LAnGuage INformation forum）ALAGIN¹では様々な言語資源を提供している [橋本 12d, 橋本 12e]．

G.1.1.1 動詞含意関係データベース

『動詞含意関係データベース』²は，含意関係が成立している動詞のペアと含意関係が成立していない動詞のペア，計 121,508 ペアを列挙したものである．ここで，動詞 1 が動詞 2 を含意する（動詞 1 → 動詞 2）とは，「動詞 1 の表す事態が成立するなら，同時かそれ以前に，動詞 2 の表す事態も成立している」ということを意味する．

正例群は次の 4 つから構成される．

- (1) 「含意が成り立つ類義／上位下位関係」（e.g. 挑戦する → チャレンジする, チンする → 加熱する）
- (2) 「文字列上包含関係にあり，含意が成り立つ類義／上位下位関係」（e.g. あざ笑う → 笑う, セリーグ優勝する → リーグ優勝する）
- (3) 「前提関係」（e.g. 酔っぱらう → 飲む, 稲刈する → 田植する）
- (4) 「作用反作用関係」（e.g. 借りる → 貸す, 受取る → 手渡す）

¹<http://www.alagin.jp/>

²<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-2>

負例群は次の4つから構成される。

- (1) 「含意，反義，予測関係ではない関連語ペア」(e.g. 通勤する \nrightarrow 走る, 読書する \nrightarrow 寛ぐ)
- (2) 「文字列上包含関係にあるが，含意，反義，予測関係ではない関連語ペア」(e.g. 冴渡る \nrightarrow 渡る, 準優勝する \nrightarrow 優勝する)
- (3) 「反義関係」(e.g. 閉める \nrightarrow 開ける, 反比例する \nrightarrow 比例する)
- (4) 「予測関係」(e.g. 紅葉する \nrightarrow 落葉する, 深煎りする \nrightarrow 挽く)

動詞含意関係データベースは，手法 [Hashimoto 09, 橋本 11a, Lin 01, Shinzato 08, Szpektor 08, Weeds 03] によって自動獲得した動詞ペアと，手法 [橋本 11b, Hashimoto 11c] で自動生成した同義句ペアから抽出した動詞ペアを統合し，人手でアノテーションしたものである。また，動詞含意関係に関して，「 $A \rightarrow B$ 」および「 $B \rightarrow C$ 」という知識から「 $A \rightarrow C$ 」となる知識を，分布類似度や含意段数の制限により適切なもののみを抽出する試みもある [橋本 10]。

このデータベースは動詞のみを対象としているが，「 X (生物)に Y (損害)を与える \rightarrow X (生物)が Y (損害)を受ける」というようにフレーズレベルでの含意関係知識を獲得する研究もある [Kloetzer 13a]。また，「 X が Y を促進する \perp X が Y を予防する」といった矛盾関係を獲得する研究もある [Kloetzer 13b, Kloetzer 14]。

なお，以下のようなフレーズレベルの含意矛盾知識のコーパスが公開されている。

述部意味関係コーパス

『述部意味関係コーパス』³は京都大学の黒橋研究室から提供されている。「名詞-格助詞-述部」ペアに対して，「同義」「含意」「反義」「無関係」のいずれかを付与したデータセットである [Izumi 14a, 泉 14b]。反義関係はさらに「経時反義」(ある状態の起点と終点)と「視点反義」(立場／視点が違う)，「属性反義」(真逆の属性を表す)に分類される。コーパスは名詞 880 個，述部ペア 10,023 個からなる。例えば次のようなデータが含まれる。

被害-ヲ-与える	被害-ヲ-もたらす	同義	
被害-ヲ-もたらす	被害-ヲ-出す	同義	
被害-ガ-甚大だ	被害-ガ-大きい	含意	
被害-ヲ-受ける	被害-ヲ-与える	反義	視点反義
被害-ガ-多い	被害-ガ-少ない	反義	属性反義
被害-ガ-拡大	被害-ヲ-免れる	無関係	

³<http://nlp.ist.i.kyoto-u.ac.jp/index.php?PredicateEvalSet>

G.1.1.2 文脈類似語データベース

『文脈類似語データベース』⁴ [Kazama 08b, 風間 09] には, 約 100 万語に対する文脈類似語 (出現文脈が類似している名詞, 最大 500 語) が類似度とともに格納されている. 例えば, 「チャイコフスキー」の文脈類似語として次のような単語が登録されている.

「チャイコフスキー」の文脈類似語

ブラームス (− 0.152), シューマン (− 0.163), メンデルスゾーン (− 0.166), ショスタコーヴィチ (− 0.178), シベリウス (− 0.18), ハイドン (− 0.181), ヘンデル (− 0.181), ラヴェル (− 0.182), シューベルト (− 0.187), ベートーヴェン (− 0.19), ドヴォルザーク (− 0.192), ラフマニノフ (− 0.193), バルトーク (− 0.198), ...

G.1.1.3 基本的意味関係の事例ベース

『基本的意味関係の事例ベース』⁵は, 文脈類似語データベースをもとに, 文脈類似度の高い 2 語間の関係を分類し, 黒田らの異表記分類基準 [黒田 10] を参考にラベル付けした結果を収録したものである. 「異表記対」「略記対」「異形同義語対」「対義語対」「部分・全体語対」「同類語対」が含まれている.

1. 異表記対

e.g. [一リーグ制, 1 リーグ制] [100 メートル, 100m] [ゆううつ, 憂鬱] [肩かけ, 肩掛け] [アタリ, ATARI]

2. 略記対

e.g. [インターネット中継, ネット中継] [短大, 短期大学] [年パス, 年間パスポート]

3. 異形同義語対

e.g. [ご飯, 食事] [ガウス分布, 正規分布] [単独トップ, 単独首位] [ワルキューレ, ヴァルキリー]

4. 対義語対

e.g. [右側, 左側] [低抵抗, 高抵抗] [インフレ, デフレ]

5. 部分・全体語対

e.g. [手, 親指] [椅子, 背もたれ] [ジョン・レノン, ビートルズ] [太平洋戦争, 第二次世界大戦]

6. 同類語対

e.g. [アロマセラピー, リフレクソロジー] [すすめ, ハト] [すべり台, ジャングルジム] [急行, 特急] [串, 割り箸]

⁴<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-1>

⁵<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-9>

G.1.1.4 日本語異表記対データベース

『日本語異表記対データベース』⁶は、文字レベルの編集距離の近い、日本語の語句の異表記対（あるいは「表記揺れの対」）の正例と負例を集めたデータである。このデータベースは人手生成データベースと自動獲得データベースの二種類からなる。人手生成データベースは、類似度の高い語句の集合から自動で生成した異表記対候補を人手でチェックしたデータであり、確実な正例として 48,067 例、確実な負例として 2,758 例、正例か負例か確実に判断できない例として 10,730 例を含む。例えば、次のようなデータが登録されている。

アクションクリエイター	アクションクリエイター	1.60129
アクションコーディネイト	アクションコーディネート	1.10272
ウォーマシン	ウォーマシーン	1.21594
世界中がアイ・ラブ・ユー	世界中がアイ・ラヴ・ユー	1.09566
世界選手権の成績	世界選手権成績	0.401025
主なみどころ	主な見どころ	1.19975
主な勝ち鞍	主な勝鞍	0.24511
仲町通	仲町通り	1.00205

一方、自動獲得データベースは、Web1 億ページに出現する頻度上位約 1,000 万の日本語語句（主として単語）に対して「編集距離 1 の異表記対」を自動獲得した結果である [小島 10]。

G.1.1.5 上位語階層データ

『上位語階層データ』⁷は、Wikipedia から自動獲得した上位下位関係の上位語を人手で階層化したものである [黒田 09]。合計約 69,000 名詞句から成る階層的シソーラスである。WordNet で得られる上位／下位関係は抽象的なものが多いのに対して、上位語階層データからはより具体的な関係が得られる。

表 G.1: 上位語階層データの例

上位語	下位語
仏像	七面大明神像
文房具	スティックのり
プログラミング言語	prolog
トルコの都市	アンカラ

⁶<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-7>

⁷<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-4>

G.1.1.6 日本語パターン言い換えデータベース

『日本語パターン言い換えデータベース』⁸は、文の係り受け解析の結果を利用して、「A が B の原因となる」というような、文内の任意の名詞 A と名詞 B を結ぶ表現の言い換えパターンを集めたデータベースである [Saeger 09b]. 各々のパターンに対して類似したパターンとその類似度が列挙されている。例えば「〈A が B の原因となる〉」という入力に対して次のような結果が得られる。

〈 A は B の原因になる 〉	0.0578512397
〈 A は B の原因 〉	0.0400322407
〈 B の原因である A 〉	0.0370898716
〈 A は B の原因にもなる 〉	0.0346598203
〈 B の原因となる A 〉	0.0335473370
...	

「〈A は B が豊富です〉」という入力に対しては次のような結果が得られる。

〈 A は B が豊富 〉	0.0549719888
〈 A には B が豊富に含まれています 〉	0.0382925298
〈 A は B も豊富です 〉	0.0377786173
〈 A は B を多く含む 〉	0.0336538462
〈 A は B も豊富 〉	0.0331325301
〈 B を豊富に含む A 〉	0.0314937013
...	

「〈A は B を防ぐ〉」という入力に対しては次のような結果が得られる。

〈 A が B を防ぐ 〉	0.0224161276
〈 A は B を予防する 〉	0.0186121788
〈 A で B を防ぐ 〉	0.0175963197
〈 B を防ぐ A 〉	0.0175141447
〈 A は B を防止する 〉	0.0132786565
〈 B を予防する A 〉	0.0132532850
...	

⁸<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-6>

G.1.2 GSK の言語資源

言語資源協会（GSK）⁹では様々な言語資源を提供している。

G.1.2.1 京都大学格フレーム

『京都大学格フレーム』¹⁰は、ウェブから収集した日本語 16 億文を用いて自動構築した格フレームからなる。格フレームとは、述語とそれが格関係をもつ語を記述したものである。約 4 万個の述語それぞれについて、平均 13 個の格フレームが構築されている。

G.1.2.2 拡張固有表現タグ付きコーパス

『拡張固有表現タグ付きコーパス』¹¹は、国立国語研究所『現代日本語書き言葉均衡コーパス』（BCCWJ）¹²のコアデータ（約 2,000 文書）および、「CD-毎日新聞’ 95 データ集」¹³の新聞記事（約 8,000 記事）に対し、関根の拡張固有表現階層-7.1.0-¹⁴を人手付与したコーパスである。

G.1.2.3 岩波国語辞典第五版タグ付きコーパス 2004

『岩波国語辞典第五版タグ付きコーパス 2004』¹⁵は、岩波国語辞典第五版における約 5 万 6 千項目のデータに、形態素・統語構造・照応と共参照、岩波国語辞典自身に基づく語義の情報などを付与したコーパスである。

G.1.2.4 新聞記事 GDA コーパス 2004

『新聞記事 GDA コーパス 2004』¹⁶は、新聞記事テキスト（3,000 記事，約 37,000 文，約 910,000 語）に対して形態素・統語構造・語義・照応と共参照の情報を付与したコーパスである。

⁹<http://www.alagin.jp/>

¹⁰<https://www.gsk.or.jp/catalog/gsk2008-b/>

¹¹<https://www.gsk.or.jp/catalog/gsk2014-a/>

¹²https://pj.ninjal.ac.jp/corpus_center/bccwj/

¹³<http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html>

¹⁴<https://sites.google.com/site/extendednamedentityhierarchy/>

¹⁵<https://www.gsk.or.jp/catalog/gsk2010-a/>

¹⁶<https://www.gsk.or.jp/catalog/gsk2009-b/>

G.1.3 その他の言語資源

G.1.3.1 EDR 電子化辞書

『EDR 電子化辞書』¹⁷は、「単語辞書」「対訳辞書」「概念辞書」「共起辞書」「専門用語辞書」「EDR コーパス」から構成される。特に、概念辞書は、41 万概念から構成される概念体系レコードの集合であり、各概念体系レコードは見出し情報、及び定義情報から構成される。これを利用することで、概念の上位／下位関係や類似度の算出が行える。

G.1.3.2 WordNet

『WordNet』¹⁸は英語の概念辞書である。WordNet 上の語は類義関係のセット (synset) でグループ化され、一つの synset が一つの概念に対応するように作られている。また、synset の結びつきにより上位／下位関係などが表されている。日本語版 WordNet¹⁹の開発も進められているが、英語版に比べて登録されている語が少なかったり、反義関係が割り当てられていないなど、英語版に比べて開発は遅れている。

G.1.3.3 日本語 WordNet 同義対データベース

『日本語 WordNet 同義対データベース』²⁰には、日本語 WordNet において、同じ synset (同じ概念を共有する語のまとまり) に掲載されている語を組み合わせで語対とし、このうち人手で同義関係にあると判定された 11,753 対が収録されている。例えば、「いざこざ」と「トラブル」のような同義ペアが収録されている。

G.1.3.4 現代日本語書き言葉均衡コーパス (BCCWJ)

『現代日本語書き言葉均衡コーパス』(BCCWJ)²¹は、現代日本語の書き言葉の全体像を把握するために構築した均衡コーパスである。書籍全般、雑誌全般、新聞、白書、ブログ、ネット掲示板、教科書、法律などのジャンルにまたがって 1 億 430 万語のデータを格納している。

¹⁷https://www2.nict.go.jp/out-promotion/techtransfer/EDR/J_index.html

¹⁸<https://wordnet.princeton.edu/>

¹⁹<http://compling.hss.ntu.edu.sg/wnja/>

²⁰<http://nlpwww.nict.go.jp/wn-ja/jpn/downloads.html>

²¹http://pj.ninjal.ac.jp/corpus_center/bccwj/

G.1.3.5 日本語ウェブコーパス 2010

『日本語ウェブコーパス 2010』²²は ipadic-2.7.0 の見出し語をシードとして, Yahoo! Web API による検索結果に含まれるウェブページを収集したものである。2010 年 6 月から 9 月にかけて収集された約 1 億ウェブページからなる。HTML アーカイブやテキストアーカイブの他, N-gram コーパスなどが提供されている。

G.1.3.6 NAIST テキストコーパス

『NAIST テキストコーパス』²³は京都テキストコーパスで利用されている毎日新聞 95 年 1 月 1 日から 17 日までの全記事, 約 2 万文, 1 月から 12 月までの社説記事, 約 2 万文, 計約 4 万文に対して, 述語と表層格 (ガ格, ヲ格, ニ格) の関係, 事態性名詞と表層格 (ガ格, ヲ格, ニ格) の関係, 事態性名詞の名詞クラス, 名詞句間の共参照関係, 指示連体詞・代名詞の照応関係の情報を付与したコーパスである。

G.1.3.7 格助詞変換データ

『格助詞変換データ』²⁴は, 京都大学テキストコーパス Version 3.0 に現れる助詞 (格助詞, 副助詞) に対してタグを付与したデータで次の三種類のデータが収録されている [飯田 10]。

(1) 使役文・受身文の能動文への変換における格助詞変換データ

受身文および使役文を能動文に変換した場合に, 元の格助詞がどのような格助詞に変換されるべきかを表すタグが付与されている。

(2) 受身文の能動文への変換における格助詞変換データ

受身文のみを対象に, それを能動文に変換した場合に, 元の格助詞がどのような格助詞に変換されるべきかを表すタグが付与されている。

(3) 格解析用データ

副助詞「は」に関して, それを格解析した場合に出力されるべき正解の格助詞を付与したデータと, 同コーパスに出現する連体修飾節の述語とその述語が修飾する名詞の格関係を解析した場合に出力されるべき正解の格助詞を付与したデータが格納されている。

²²<http://s-yata.jp/corpus/nwc2010/>

²³<http://chasen.naist.jp/hiki/naistcorpus/>

²⁴<https://alaginrc.nict.go.jp/case/index.html>

G.1.3.8 「拡張固有表現+ Wikipedia」データ

「拡張固有表現+ Wikipedia」データ²⁵は、Wikipedia の各項目に拡張固有表現²⁶のタグを振ったデータである [関根 18a, 関根 18b].

G.1.3.9 Wikipedia 記事への促進・抑制関係付与コーパス

『Wikipedia 記事への促進・抑制関係付与コーパス』²⁷は、Wikipedia の記事, 1,494 件に対して, 促進・抑制関係を付与したコーパスである [埴 17]. 記事の概要に含まれる表現に対して, 以下の 4 種類のラベルがアノテーションされている. 記事のタイトルが促進するもの (PRO), タイトルが抑制するもの (SUP), タイトルを促進するもの (PRO_BY), タイトルを抑制するもの (SUP_BY) .

G.1.3.10 解答可能性付き読解データセット

『解答可能性付き読解データセット』²⁸は, 約 5 万 5 千件の質問・解答・文書の組に対して, 文書に質問の正解の根拠が書かれているかどうかを, クラウドソーシングを利用して, 人手で判定したデータセットである [鈴木 18].

G.1.3.11 Twitter 日本語評判分析データセット

『Twitter 日本語評判分析データセット』²⁹は, 携帯電話などを話題にした, 約 50 万件のツイートに対してクラウドソーシングを利用して, ポジティブ/ネガティブ/ニュートラルを判定したデータセットである [芥子 17].

G.1.3.12 評価対象-評価表現抽出用 日本語 Twitter データセット

『評価対象-評価表現抽出用 日本語 Twitter データセット』³⁰は, ツイートに評価情報が含まれるか否かを付与したデータと, 評価対象と評価表現を表す部分文字列をアノテーションした

²⁵<http://www.languagecraft.com/enew/>

²⁶<https://sites.google.com/site/extendednamedentity711/>

²⁷http://www.cl.ecei.tohoku.ac.jp/wikipedia_pro_sup/

²⁸<http://www.cl.ecei.tohoku.ac.jp/rcqa/>

²⁹<http://bigdata.naist.jp/~ysuzuki/data/twitter/>

³⁰http://www.cl.ecei.tohoku.ac.jp/resources/twitter_target_review/

データから構成される [栗原 18].

G.1.3.13 単語感情極性対応表

単語感情極性対応表³¹は、高村らがスピンモデルを用いて単語の感情極性を自動で計算し [Takamura 05a, Takamura 05a, Takamura 06], -1 (ネガティブ) から+1 (ポジティブ) の実数値を単語に付与された辞書である。例えば、「楽しい」は 0.995837 であり、「苦しい」は -0.999788 である。

G.1.3.14 日本語評価極性辞書

日本語評価極性辞書³²は、用言に関する辞書と名詞に関する辞書がある。用言に関する辞書は、約 5000 表現からなり、各表現を客観的／主観的に分類したうえで、ポジティブ／ネガティブに分類している [小林 05]。名詞に関する辞書は、約 8500 表現からなり、各名詞に「～になる」「～がある」などの述語を付与したときに、ポジティブ／ネガティブ／ニュートラルであるか分類した [東山 08]。

G.1.3.15 Polar Phrase Dictionary

『Polar Phrase Dictionary』³³は、HTML 文書から評価を含む文を自動で抽出し、作成された評価文コーパス [Kaji 06] から、形容詞または形容詞句に関する評価極性値を自動で計算し作成された評価表現辞書である [Kaji 07]。例えば、「心配が無い」は 15.76368708 であり、「駅から遠い」は -14.32184005 である。

G.1.3.16 日本語感情表現辞書

『日本語感情表現辞書』³⁴は、約 2,000 個の表現について、48 種類の感情を付与した辞書である。例えば、「言い寄る」は親しみであり、「唾み合う」は嫌悪である。

³¹http://www.lr.pi.titech.ac.jp/%7Etakamura/pndic_ja.html

³²<http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary>

³³<http://www.tkl.iis.u-tokyo.ac.jp/~kaji/polardic/>

³⁴<http://www.jnlp.org/SNOW/D18>

G.1.3.17 やさしい日本語コーパス

『やさしい日本語コーパス』³⁵は、日英対訳コーパス³⁶の日本語について、やさしい日本語に書き換えた5万文からなる対訳コーパスである [Maruyama 18, 山本 17].

G.1.3.18 Simple PPDB: Japanese

『Simple PPDB: Japanese』³⁷は、57万単語について『日本語教育語彙表』に基づく難易度(「初級」「中級」「上級」)を付与した『単語難易度辞書』と、この辞書と日本語言い換えデータベース『PPDB: Japanese』³⁸から作成した『平易な言い換え辞書』から構成される [梶原 17].

G.1.3.19 日本語教育語彙表

『日本語教育語彙表』³⁹は、約18000語の日本語教育用の語彙に対して、初級前半、初級後半、中級前半、中級後半、上級前半、上級後半を付与したデータベースである [砂川 12a, Sunakawa 12b]. 例えば、「食事」は初級後半であり、「飢餓」は上級前半である.

G.1.3.20 基本語データベース 語義別親密度

『基本語データベース 語義別親密度』⁴⁰は、約2.8万単語、約4.5万語義に対して、1(親密度が低い)から7(親密度が高い)の範囲で親密度を付与したデータベースである. 例えば、「縊れる」は1.175と親密度が低く、「働く」は6.3と親密度が高い.

³⁵<http://www.jnlp.org/SNOW/T15>

³⁶https://github.com/odashi/small_parallel_enja

³⁷<https://github.com/tmu-nlp/simple-jppdb>

³⁸<https://ahcweb01.naist.jp/resource/jppdb/>

³⁹<http://jhlee.sakura.ne.jp/JEV.html>

⁴⁰https://hon.gakken.jp/reference/special/jiten/kihongo_db/index.html

G.2 言語処理ツール

G.2.1 形態素解析器

G.2.1.1 JUMAN / JUMAN++

JUMAN⁴¹は、黒橋・河原研究室で開発されている日本語形態素解析器である。解析結果として、品詞⁴²や活用形、読みに関する情報の他、カテゴリやドメインに関する情報も付与される(図 G.1, 図 G.2)。RNN 言語モデルを用いることで、広い文脈での意味的な整合性を改善した JUMAN++⁴³が開発されている [Morita 15, 森田 16, Tolmachev 18a, Tolmachev 18b]。

カテゴリ名	例	カテゴリ名	例
人	学生, 先生, …	場所-施設	ビル, 公園, …
組織・団体	政府, 企業, …	場所-施設部位	天井, 床, …
動物	犬, 猫, …	場所-自然	山, 海, …
植物	桜, バラ, …	場所-機能	上, 下, …
動物-部位	手, 足, …	場所-その他	都市, 村, …
植物-部位	葉, 枝, …	抽象物	思考, 理由, …
人工物-食べ物	パン, コーヒー, …	形・模様	円, 球, …
人工物-衣類	ズボン, セーター, …	色	赤, 青, …
人工物-乗り物	自動車, 飛行機, …	数量	複数, メートル, …
人工物-金銭	給料, 借金, …	時間	今日, 朝, …
人工物-その他	鉛筆, 消しゴム, …		
自然物	石, 岩, …		

図 G.1: JUMAN のカテゴリ情報

ドメイン名	例	ドメイン名	例
文化・芸術	映画 音楽 御輿 …	交通	駅 道路 アクセル …
レクリエーション	観光 花火 カジノ …	教育・学習	先生 算数 開校 …
スポーツ	選手 野球 角界 …	科学・技術	研究 理論 ウラン …
健康・医学	手術 診断 胃液 …	ビジネス	輸入 市場 売上 …
家庭・暮らし	育児 家具 帰省 …	メディア	放送 記者 載る …
料理・食事	箸 昼食 和える …	政治	司法 税 挙党 …
ドメイン無し	青 感情 上がる …		

図 G.2: JUMAN のドメイン情報

⁴¹<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

⁴²<http://www.unixuser.org/~euske/doc/postag/#juman>

⁴³<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

「羽生結弦が国際大会を欠場した」を JUMAN で解析した結果を以下に示す.

```
羽生 はぶ 羽生 名詞 6 人名 5 * 0 * 0 "人名:日本:姓:634:0.00043"
@ 羽生 はにゅう 羽生 名詞 6 地名 4 * 0 * 0 "代表表記:羽生/はにゅう 地名:日本:埼玉県:市"
結 けつ 結 名詞 6 普通名詞 1 * 0 * 0 "代表表記:結/けつ 漢字読み:音 カテゴリ:抽象物"
弦 げん 弦 名詞 6 普通名詞 1 * 0 * 0 "代表表記:弦/げん 漢字読み:音 カテゴリ:人工物-その他"
が が が 助詞 9 格助詞 1 * 0 * 0 NIL
国際 こくさい 国際 名詞 6 普通名詞 1 * 0 * 0 "代表表記:国際/こくさい カテゴリ:抽象物"
大会 たいかい 大会 名詞 6 普通名詞 1 * 0 * 0 "代表表記:大会/たいかい カテゴリ:抽象物"
を を を 助詞 9 格助詞 1 * 0 * 0 NIL
欠場 けつじょう 欠場 名詞 6 サ変名詞 2 * 0 * 0 "代表表記:欠場/けつじょう カテゴリ:抽象物
ドメイン:スポーツ 反義:名詞-サ変名詞:休場/きゅうじょう; 名詞-サ変名詞:出場/しゅつじょう"
した した する 動詞 2 * 0 サ変動詞 16 タ形 10 "代表表記:する/する 付属動詞候補 (基本) 自
他動詞:自:成る/なる"
EOS
```

G.2.1.2 茶筌 (ChaSen)

ChaSen⁴⁴は奈良先端科学技術大学院大学の松本研究室で開発されている. ChaSen ではHMMに基づいて接続コストや単語生起コストを推定している. ChaSen の品詞体系は任意の階層化を許しており, JUMAN よりも品詞の分類が細かい⁴⁵.

「羽生結弦が国際大会を欠場した」を ChaSen で解析した結果を以下に示す.

```
羽生 ハブ 羽生 名詞-固有名詞-人名-姓
結 ユイ 結 名詞-一般
弦 ツル 弦 名詞-一般
が ガ が 助詞-格助詞-一般
国際 コクサイ 国際 名詞-一般
大会 タイカイ 大会 名詞-一般
を ヲ を 助詞-格助詞-一般
欠場 ケツジョウ 欠場 名詞-サ変接続
した シタ した 名詞-一般
EOS
```

⁴⁴<http://chasen-legacy.osdn.jp/>

⁴⁵<http://www.unixuser.org/~euske/doc/postag/#chasen>

G.2.1.3 MeCab

MeCab⁴⁶は京都大学情報学研究科を通じて開発されたオープンソースの形態素解析エンジンである。MeCab は言語, 辞書, コーパスに依存しない汎用的な設計を基本方針としており, 辞書としては, IPA 辞書⁴⁷, NAIST 辞書⁴⁸, UniDic⁴⁹, JUMAN 辞書⁵⁰, NEologd⁵¹などが利用できる。また, パラメータの推定に CRF を用いており, ChaSen が採用している HMM に比べ解析性能は高い [Kudo 04]。さらに, 平均的に ChaSen や JUMAN に比べて高速に動作する。

「羽生結弦が国際大会を欠場した」を MeCab で解析した結果を以下に示す。

```
羽生 名詞,固有名詞,人名,姓,*,*,羽生,ハブ,ハブ,,
結 名詞,一般,*,*,*,結,ユイ,ユイ,,
弦 名詞,一般,*,*,*,弦,ツル,ツル,,
が 助詞,格助詞,一般,*,*,*,が,ガ,ガ,,
国際 名詞,一般,*,*,*,国際,コクサイ,コクサイ,,
大会 名詞,一般,*,*,*,大会,タイカイ,タイカイ,,
を 助詞,格助詞,一般,*,*,*,を,ヲ,ヲ,,
欠場 名詞,サ変接続,*,*,*,欠場,ケツジョウ,ケツジョー,,
し 動詞,自立,*,*,サ変・スル,連用形,する,シ,シ,,
た 助動詞,*,*,特殊・タ,基本形,た,タ,タ,,
EOS
```

G.2.2 係り受け解析器

G.2.2.1 KNP

KNP⁵²は, 自動構築した大規模格フレームに基づく統合的確率モデルにより構文解析を行うシステムである [Kawahara 06b, 河原 07]。大規模な格フレームに基づくゼロ照応解析 [Sasano 11a, 笹野 11b], CRF に基づく固有表現解析, ルールに基づく共参照解析など多様な解析結果を出力可能である [笹野 13]。JUMAN や JUMAN++ の形態素解析結果を入力として与える。

⁴⁶<http://taku910.github.io/mecab/>

⁴⁷<http://sourceforge.net/projects/mecab/files/>

⁴⁸<http://sourceforge.jp/projects/naist-jdic/releases/>

⁴⁹<https://ja.osdn.net/projects/unidic/>

⁵⁰<http://sourceforge.net/projects/mecab/files/>

⁵¹<https://github.com/neologd/mecab-ipadic-neologd>

⁵²<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

「羽生結弦が国際大会を欠場した」を KNP で解析した結果を以下に示す。

```
* 2D<文頭><人名><ガ><助詞><体言><係:ガ格><区切:0-0><格要素><連用要素><正規化代表表記:羽生/はぶ?羽生/はにゅう+結弦/けつげん><主辞代表表記:結弦/けつげん>
+ 1D<文節内><係:文節内><文頭><人名疑><地名疑><体言><名詞項候補><先行詞候補><正規化代表表記:羽生/はぶ?羽生/はにゅう><NE内:PERSON><EID:0>
羽生 はぶ 羽生 名詞 6 人名 5 * 0 * 0 "人名:日本:姓:634:0.00043 疑似代表表記 代表表記:羽生/はぶ" (人名:日本:姓:634:0.00043) <疑似代表表記> <代表表記:羽生/はぶ> <正規化代表表記:羽生/はぶ?羽生/はにゅう> <品曖> <ALT-羽生-はにゅう-羽生-6-4-0-0">
代表表記:羽生/はにゅう 地名:日本:埼玉県:市" <地名:日本:埼玉県:市> <品曖-人名> <品曖-地名> <原形曖昧> <文頭> <漢字>
<かな漢字> <名詞相当語> <自立> <内容語> <タグ単位始> <文節始> <固有キー> <NE:PERSON:B>
+ 4D<人名><ガ><助詞><体言><係:ガ格><区切:0-0><格要素><連用要素><名詞項候補><先行詞候補><SM-人>
<SM-主体> <正規化代表表記:結弦/けつげん> <NE:PERSON:羽生結弦> <照応詞候補:羽生結弦> <解析格:ガ> <EID:1>
結弦 けつげん 結弦 名詞 6 人名 5 * 0 * 0 "代表表記:結弦/けつげん 漢字読み:音 カテゴリ:抽象物 品詞変更:結-けつ-結-6-1-0-0"
代表表記:結弦/けつげん <漢字読み:音> <カテゴリ:抽象物> <正規化代表表記:結弦/けつげん> <形態素連結-人名> <人名:日本:名:自動認識> <漢字> <かな漢字> <名詞相当語> <自立> <複合<> <内容語> <タグ単位始> <固有キー> <文節主辞> <NE:PERSON:E>
が が 助詞 9 格助詞 1 * 0 * 0 NIL <かな漢字> <ひらがな> <付属>
* 2D<ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><正規化代表表記:国際/こくさい+大会/たいかい>
<主辞代表表記:大会/たいかい>
+ 3D<文節内><係:文節内><体言><名詞項候補><先行詞候補><正規化代表表記:国際/こくさい> <照応詞候補:国際> <EID:2>
国際 こくさい 国際 名詞 6 普通名詞 1 * 0 * 0 "代表表記:国際/こくさい カテゴリ:抽象物" <代表表記:国際/こくさい> <カテゴリ:抽象物> <正規化代表表記:国際/こくさい> <漢字> <かな漢字> <名詞相当語> <自立> <内容語> <タグ単位始> <文節始>
+ 4D<ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><名詞項候補><先行詞候補><正規化代表表記:大会/たいかい> <照応詞候補:国際大会> <解析格:ヲ> <EID:3> <述語項構造:国際/こくさい+大会/たいかい:名 1>
大会 たいかい 大会 名詞 6 普通名詞 1 * 0 * 0 "代表表記:大会/たいかい カテゴリ:抽象物" <代表表記:大会/たいかい> <カテゴリ:抽象物> <正規化代表表記:大会/たいかい> <漢字> <かな漢字> <名詞相当語> <自立> <複合<> <内容語> <タグ単位始>
<文節主辞>
を を 助詞 9 格助詞 1 * 0 * 0 NIL <かな漢字> <ひらがな> <付属>
* -1D<文末><サ変><サ変動詞><時制-過去><用言:動><レベル:C><区切:5-5><ID:(文末)><提題受:30><主節><動態述語>
<正規化代表表記:欠場/けつじょう> <主辞代表表記:欠場/けつじょう>
+ -1D<文末><サ変動詞><時制-過去><用言:動><レベル:C><区切:5-5><ID:(文末)><提題受:30><主節><動態述語>
<サ変> <正規化代表表記:欠場/けつじょう> <用言代表表記:欠場/けつじょう> <主題格:一人称優位> <格関係 1:ガ:結弦>
<格関係 3:ヲ:大会> <格解析結果:欠場/けつじょう:動 1:ガ/C/結弦/1/0/1; ヲ/C/大会/3/0/1; ニ/U/-/-/-/-; ト/U/-/-/-/-;
デ/U/-/-/-/-; カラ/U/-/-/-/-; マデ/U/-/-/-/-; ヘ/U/-/-/-/-; 時間/U/-/-/-/-; 外の関係/U/-/-/-/-; 修飾/U/-/-/-/-; ノ/U/-/-/-/-;
ニヨル/U/-/-/-/-; ニツヅク/U/-/-/-/-; トスル/U/-/-/-/-; ヲフクメル/U/-/-/-/-; ニトモナウ/U/-/-/-/-> <EID:4> <述語項構造:欠場/けつじょう:動 1:ガ/C/羽生結弦/1; ヲ/C/国際大会/3>
欠場 けつじょう 欠場 名詞 6 サ変名詞 2 * 0 * 0 "代表表記:欠場/けつじょう カテゴリ:抽象物 ドメイン:スポーツ 反義:名詞-サ変名詞:休場/きゅうじょう; 名詞-サ変名詞:出場/しゅつじょう" <代表表記:欠場/けつじょう> <カテゴリ:抽象物> <ドメイン:スポーツ>
<反義:名詞-サ変名詞:休場/きゅうじょう; 名詞-サ変名詞:出場/しゅつじょう> <正規化代表表記:欠場/けつじょう> <漢字>
<かな漢字> <名詞相当語> <サ変> <サ変動詞> <自立> <内容語> <タグ単位始> <文節始> <文節主辞>
した した する 動詞 2 * 0 サ変動詞 16 タ形 10 "代表表記:する/する 付属動詞候補 (基本) 自他動詞:自:成る/なる" <代表表記:する/する> <付属動詞候補 (基本)> <自他動詞:自:成る/なる> <正規化代表表記:する/する> <文末> <表現文末> <と系連用テ形複合辞> <かな漢字> <ひらがな> <活用語> <付属>
EOS
```

G.2.2.2 CaboCha

CaboCha⁵³はSVMに基づく日本語係り受け解析器である [Kudo 00, Kudo 02a, 工藤 02b]. CaboChaの特徴として, IREXの定義による固有表現解析が可能なこと, 係り受けの同定に使用する素性をユーザ側で再定義可能なこと, 内部の辞書に高速なTrie構造であるDouble-Arrayを採用していることなどが挙げられる.

⁵³<http://taku910.github.io/cabocha/>

「羽生結弦が国際大会を欠場した」を CaboCha で解析した結果を以下に示す。

```
羽生 名詞, 固有名詞, 人名, 姓, *, *, 羽生, ハブ, ハブ,, B-PERSON
結 名詞, 一般, *, *, *, 結, ユイ, ユイ,, I-PERSON
弦 名詞, 一般, *, *, *, 弦, ツル, ツル,, I-PERSON
が 助詞, 格助詞, 一般, *, *, *, が, ガ, ガ,, O
* 1 2D 1/2 -1.929536
国際 名詞, 一般, *, *, *, 国際, コクサイ, コクサイ,, O
大会 名詞, 一般, *, *, *, 大会, タイカイ, タイカイ,, O
を 助詞, 格助詞, 一般, *, *, *, を, ヲ, ヲ,, O
* 2 -1D 1/2 0.000000
欠場 名詞, サ変接続, *, *, *, 欠場, ケツジョウ, ケツジョー,, O
し 動詞, 自立, *, *, サ変・スル, 連用形, する, シ, シ,, O
た 助動詞, *, *, *, 特殊・タ, 基本形, た, タ, タ,, O
EOS
```

G.2.2.3 J.DepP

J.DepP (Japanese Dependency Parsers)⁵⁴は、線形分類器を用いた高速な係り受け解析器である [Yoshinaga 12, Yoshinaga 14]。JUMAN や MeCab の形態素解析結果を入力として与える。

「羽生結弦が国際大会を欠場した」を J.DepP で解析した結果を以下に示す。

```
* 0 1D
羽生 はぶ 羽生 名詞 6 人名 5 * 0 * 0 "人名:日本:姓:634:0.00043"
@ 羽生 はにゅう 羽生 名詞 6 地名 4 * 0 * 0 "代表表記:羽生/はにゅう 地名:日本:埼玉県:市"
* 1 3D
結 けつ 結 名詞 6 普通名詞 1 * 0 * 0 "代表表記:結/けつ 漢字読み:音 カテゴリ:抽象物"
弦 げん 弦 名詞 6 普通名詞 1 * 0 * 0 "代表表記:弦/げん 漢字読み:音 カテゴリ:人工物-その他"
が が が 助詞 9 格助詞 1 * 0 * 0 NIL
* 2 3D
国際 こくさい 国際 名詞 6 普通名詞 1 * 0 * 0 "代表表記:国際/こくさい カテゴリ:抽象物"
大会 たいかい 大会 名詞 6 普通名詞 1 * 0 * 0 "代表表記:大会/たいかい カテゴリ:抽象物"
を を を 助詞 9 格助詞 1 * 0 * 0 NIL
* 3 -1D
欠場 けつじょう 欠場 名詞 6 サ変名詞 2 * 0 * 0 "代表表記:欠場/けつじょう カテゴリ:抽象物
ドメイン:スポーツ 反義:名詞-サ変名詞:休場/きゅうじょう; 名詞-サ変名詞:出場/しゅつじょう"
した した する 動詞 2 * 0 サ変動詞 16 タ形 10 "代表表記:する/する 付属動詞候補 (基本) 自
他動詞:自:成る/なる"
EOS
```

⁵⁴<http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp/>

G.2.3 述語項構造解析器

G.2.3.1 SynCha

SynCha (新茶)⁵⁵は、日本語の述語項構造解析器である [Iida 11]。CaboCha で求まる文の係り受け関係に基づいて、述語を中心とした文の構成要素の構造を解析する。文章中に出現する述語とその格要素、および、名詞句間の共参照関係を同定する。

「羽生結弦が国際大会を欠場した」を SynCha で解析した結果を以下に示す。

```
* 0 2D 2/3
羽生 名詞, 固有名詞, 人名, 姓,,, 羽生, ハブ, ハブ,, B-PERSON
結 名詞, 一般,* ,*, 結, ユイ, ユイ,, I-PERSON
弦 名詞, 一般,* ,*, 弦, ツル, ツル,, I-PERSON id="1"
が 助詞, 格助詞, 一般,* ,*, が, ガ, ガ,, O
* 1 2D 1/2 国際 名詞, 一般,* ,*, 国際, コクサイ, コクサイ,, O
大会 名詞, 一般,* ,*, 大会, タイカイ, タイカイ,, O id="2"
を 助詞, 格助詞, 一般,* ,*, を, ヲ, ヲ,, O
* 2 -1D 1/2 0.000000
欠場 名詞, サ変接続,* ,*, 欠場, ケツジョウ, ケツジョー,, O
し 動詞, 自立,* ,*, サ変・スル, 連用形, する, シ, シ,, O ga="1" o="2" type="pred"
た 助動詞,* ,*, 特殊・タ, 基本形, た, タ, タ,, O
EOS
EOT
```

G.2.3.2 ChaPAS

ChaPAS⁵⁶は Java ベースの日本語述語項構造解析器である。ChaPAS の特徴は、様々な言語資源を活用した解析ができること、並列構造解析機能があることなどが挙げられる。

G.2.4 その他の言語処理ツール

G.2.4.1 Zunda

Zunda⁵⁷は東北大学の乾・岡崎研究室で開発された日本語拡張モダリティ解析器である [江口 11, Matsuyoshi 10, 松吉 11]。文中のイベント（動詞や形容詞など）に対して、その真偽判断（イベントが起こったかどうか）、仮想性（仮定の話かどうか）などを解析する。文の解析には CaboCha が使われ、出力には、態度表明者、時制、仮想、態度、真偽判断、価値判断の 6 種類の情報が付与される。

⁵⁵<https://sites.google.com/site/ryuiida/syncha>

⁵⁶<https://github.com/yotarow/chapas>

⁵⁷<https://jmizuno.github.io/zunda/>

表 G.2: Zunda の出力ラベル

態度表明者	wr:筆者のみ
時制	未来, 非未来
仮想	条件, 帰結, 0
態度	叙述, 意志, 欲求, 働きかけ-直接, 働きかけ-間接, 働きかけ-勧誘, 欲求, 問いかけ
真偽判断	成立, 不成立, 不成立から成立, 成立から不成立, 高確率, 低確率, 低確率から高確率, 高確率から低確率, 0
価値判断	ポジティブ, ネガティブ, 0

「羽生結弦が国際大会を欠場した」を Zunda で解析した結果を以下に示す。

```
#EVENT0 7 wr:筆者 非未来 0 叙述 成立 0 0
#EVENT1 8 wr:筆者 非未来 0 叙述 成立 0 0
* 0 2D 2/3 -1.929536
羽生 名詞, 固有名詞, 人名, 姓, *, *, 羽生, ハブ, ハブ,,
結 名詞, 一般, *, *, *, 結, ユイ, ユイ,,
弦 名詞, 一般, *, *, *, 弦, ツル, ツル,,
が 助詞, 格助詞, 一般, *, *, *, が, ガ, ガ,,
* 1 2D 1/2 -1.929536
国際 名詞, 一般, *, *, *, 国際, コクサイ, コクサイ,,
大会 名詞, 一般, *, *, *, 大会, タイカイ, タイカイ,,
を 助詞, 格助詞, 一般, *, *, *, を, ヲ, ヲ,,
* 2 -1D 1/2 0.000000
欠場 名詞, サ変接続, *, *, *, 欠場, ケツジョウ, ケツジョー,,
し 動詞, 自立, *, *, サ変・スル, 連用形, する, シ, シ,,
た 助動詞, *, *, *, 特殊・タ, 基本形, た, タ, タ,,
EOS
```

G.2.4.2 ASA

ASA (Argument Structure Analyzer)⁵⁸は岡山大学の竹内研究室で開発されている意味役割付与システムである [竹内 10]。係り受け解析に CaboCha を使用しており、その結果に対して述語シソーラス（動詞, 形容詞, 形容動詞）に基づく意味役割と述語の概念フレームが自動で付与される。例えば、「行く」の意味役割として「状態変化あり-位置変化-位置変化（物理）-着点への移動-」が付与され、「綺麗だ」の意味役割として「状態変化なし（状態）-対象の性質-物理的状态-清／汚-清」が付与される。

⁵⁸<http://www.cl.cs.okayama-u.ac.jp/study/project/asa/>

G.2.4.3 normalizeNumexp

normalizeNumexp⁵⁹は、日本語の文中に含まれる数量表現や時間表現を抽出し、正規化するツールである [Narisawa 13].

G.2.4.4 意見(評価表現)抽出ツール

意見(評価表現)抽出ツール⁶⁰は、独立行政法人情報通信研究機構が開発し、高度言語情報融合フォーラム(ALAGIN)からオープンソースソフトウェアとして配布されている。1行1文のテキストファイルに対して、機械学習により以下の解析を行う [Nakagawa 08, 中川 09, Nakagawa 10].

- (1) 評価を表す表現の抽出
- (2) 評価の意味的な分類 (e.g. 感情, 批評, メリット)
- (3) 評価が肯定的なニュアンスを表すのか, 否定的なニュアンスを表すのかの判定
- (4) 評価を発信する主体の抽出

G.2.4.5 RaSC

RaSC (Rapid Service Connector)⁶¹は、情報通信研究機構が開発したユーザプログラムを高速・高並列に実行可能にするフリーのミドルウェアである [田仲 14]. RaSCは既存の形態素解析器や係り受け解析器などのプログラムを大規模な自然言語処理に適用することを念頭に開発されたものであり、ユーザプログラムを複数のCPUコアを活用して並列実行したり、複数の計算機上で分散実行することを容易にする。RaSC上で稼働するユーザプログラムのプロセスは、一度起動されると計算機に常駐するため、辞書ファイルをロードする言語処理プログラムのように巨大ファイルのロードなどにより起動時間が長くなっているプログラムであっても効率的に実行できる。JUMANやKNP, J.DepPなどのツールをRaSC上で実行することで、並列処理が可能となる。

⁵⁹<http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FnormalizeNumexp>

⁶⁰<https://alaginrc.nict.go.jp/opinion/>

⁶¹https://alaginrc.nict.go.jp/rasc/rasc_overview.html

謝辞

本研究を進めるにあたり，ご指導してくださった指導教員の小林哲則教授に感謝いたします．小林先生には研究内容に関する指導の他にも，プレゼンの組み方や研究者としてのあり方など，たくさんのご指導ご鞭撻いただけたことに重ねて感謝の意を示したいと思います．また，林先生，藤江先生を初め，研究に関して助言をくださった教員および学生の皆様，どうもありがとうございました．最後に，経済的にも精神的にも支えてくださった家族に感謝を申し上げます．

参考文献

- [Aberdeen 95] Aberdeen, J., Burger, J., Day, D., Hirschman, L., Robinson, P., and Vilain, M.: MITRE: Description of the alembic system used for MUC-6, in *Proceedings of the 6th Message Understanding Conference*, pp. 141-155 (1995)
- [Adigwe 18] Adigwe, A., Tits, N., Haddad, K.E., Ostadabbas, S., and Dutoit, T.: The emotional voices database: Towards controlling the emotion dimension in voice generation systems, in *Proceedings of the 6th International Conference on Statistical Language and Speech Processing* (2018)
- [Adriaens 95] Adriaens, G.: Simplified English grammar and style correction in an MT framework: the LRE SECC project, in *Aslib Proceedings*, Vol. 47, pp. 73-82 (1995)
- [Al-Sabahi 18] Al-Sabahi, K., Zuping, Z., and Nadher, M.: A hierarchical structured self-attentive model for extractive document summarization (HSSAS), *IEEE Access*, Vol. 6, pp. 24205-24212 (2018)
- [Ando 15] Ando, A., Asami, T., Okamoto, M., Masataki, H., and Sakauchi, S.: Agreement and disagreement utterance detection in conversational speech by extracting and integrating local features, in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, pp. 2494-2498 (2015)
- [Appelt 93] Appelt, D.E., Hobbs, J.R., Bear, J., Israel, D., and Tyson, M.: FASTUS: A finite-state processor for information extraction from real-world text, in *Proceedings of the 13th International Joint Conferences on Artificial Intelligence*, pp. 1172-1178 (1993)
- [稲垣 13] 稲垣博人, 長谷川隆明, 高橋敏, 松尾義博: 知的な音声言語インタフェース技術に向けた取り組み, *NTT 技術ジャーナル*, Vol. 20, No. 4, pp. 12-15 (2013)
- [Asghar 16] Asghar, N.: Automatic Extraction of Causal Relations from Natural Language Texts: A Comprehensive Survey, *arXiv:1605.07895*, pp. 1-10 (2016)

- [Ba 16] Ba, J.L., Kiros, J.R., and Hinton, G.E.: Layer normalization, *arXiv:1607.06450*, pp. 1-14 (2016)
- [Bahdanau 15] Bahdanau, D., and Cho, K., and Bengio, Y.: Neural machine translation by jointly learning to align and translate, in *Proceedings of the 3th International Conference on Learning Representations*, pp. 1-15 (2015)
- [Barrios 15] Barrios, F., Lopez, F., Argerich, L., Wachenchauser, R.: Variations of the similarity function of TextRank for automated summarization, in *Proceedings of the Argentine Symposium on Artificial Intelligence*, pp. 65-72 (2015)
- [Barzilay 03] Barzilay, R. and Elhadad, N.: Sentence alignment for monolingual comparable corpora, in *Proceedings of the 2003 conference on Empirical Methods in Natural Language Processing*, pp. 25-32 (2003)
- [Belder 10] Belder, J.D., Deschacht, K., and Moens, M.F.: Lexical simplification, in *Proceedings of the 1st International Conference on Interdisciplinary Research on Technology, Education and Communication*, pp. 1-4 (2010)
- [Belder 12] Belder, J.D. and Moens, M.F.: A dataset for the evaluation of lexical simplification, *Proceedings of the 13th international conference on Computational Linguistics and Intelligent Text Processing*, pp. 426-437 (2012)
- [別所 12] 別所史浩, 原田達也, 國吉康夫: リアルタイムクラウドソーシングと Twitter 大規模コーパスを利用した対話システム, 情報処理学会研究報告, Vol. 2012-NL-206, No. 13, pp. 1-8 (2012)
- [Biran 11] Biran, O., Brody, S., and Elhadad, N.: Putting it simply: a context-aware approach to lexical simplification, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 496-501 (2011)
- [Blair-Goldensohn 03a] Blair-Goldensohn, S., McKeown, K.R., and Schlaikjer, A.H.: A hybrid approach for QA track definitional questions, in *Proceedings of the 12th Text REtrieval Conference*, pp. 1-8 (2003)
- [Blair-Goldensohn 03b] Blair-Goldensohn, S., McKeown, K.R., and Schlaikjer, A.H.: A Hybrid Approach for Answering Definitional Questions, in *Proceedings of the 26th ACM SIGIR Conference*, pp. 1-8 (2003)

- [Blair-Goldensohn 04] Blair-Goldensohn, S., McKeown, K.R., Schlaikjer, A.H.: Answering definitional questions: A hybrid approach, *New Directions in Question Answering*, pp. 47-58 (2004)
- [Blake 07] Blake, C., Kampov, J., Orphanides, A.K., West, D., and Lown, C.: Query expansion, lexical simplification and sentence selection strategies for multi-document summarization, *Document Understanding Conference*, pp. 1-8 (2007)
- [Blum 78] Blum, S. and Levenston, E.A.: Universals of lexical simplification, *Language Learning*, Vol. 28, No. 2, pp. 399-415 (1978)
- [Bojanowski 17] Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T.: Enriching word vectors with subword information, *Transactions of the Association for Computational Linguistics*, Vol. 5, pp. 135-146 (2017)
- [Bott 11] Bott, S. and Saggion, H.: An unsupervised alignment algorithm for text simplification corpus construction, in *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pp. 20-26 (2011)
- [Bott 12] Bott, S., Rello, L., Drndarevic, B., and Saggion, H.: Can Spanish be simpler? LexSiS: Lexical simplification for Spanish, in *Proceedings of the 24th International Conference on Computational Linguistics*, pp. 357-374 (2012)
- [Brandow 95] Brandow, R., Mitze, K., and Rau, L.F.: Automatic condensation of electronic publications by sentence selection, *Information Processing and Management*, Vol. 31, No. 5, pp. 675-685 (1995)
- [Brown 84] Brown, A.L., Palincsar, A.S., and Armbruster, B.B.: Instructing comprehension-fostering activities in interactive learning situations, *Learning and Comprehension of Text*, pp. 255-286 (1984)
- [Burkhardt 05] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F., and Weiss, B.: A database of German emotional speech, in *Proceedings of the 9th European Conference on Speech Communication and Technology*, pp. 1517-1520 (2005)
- [Carbonell 98] Carbonell, J. and Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries, in *Proceedings of the 21th Annual Inter-*

- national ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 335-336 (1998)
- [Carroll 05] Carroll, J.C., Czuba, K., Duboue, P., and Prager, J.: IBM' s PIQUANT II in TREC2005, in the *Proceedings of the 14th Text REtrieval Conference*, pp. 1-8 (2005)
- [Case 07] Case, D.O.: *Looking for Information, A Survey of Research on Information Seeking, Needs and Behavior*, Second Edition, Academic Press (2007)
- [Celikyilmaz 18] Celikyilmaz, A., Bosselut, A., He, X., and Choi, Y.: Deep communicating agents for abstractive summarization, in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1662-1675 (2018)
- [Chafe 94] Chafe, W.: *Discourse, consciousness, and time*, University Of Chicago, Vol. 2 (1994)
- [Chan 05] Chan, K. and Lam, W.: Extracting causation knowledge from natural language texts, *International Journal of Intelligent System*, Vol. 20, No. 3, pp. 327-358 (2005)
- [Chandrasekar 97] Chandrasekar, R. and Srinivas, B.: Automatic induction of rules for text simplification, *Knowledge-Based Systems*, Vol. 10, No. 3, pp. 183-190 (1997)
- [Chang 04] Chang, D.S. and Choi, K.S.: Causal relation extraction using cue phrase and lexical pair probabilities, in *Proceedings of the 1st International Joint Conference on Natural Language Processing*, pp. 61-70 (2004)
- [Cheng 16] Cheng, J. and Lapata, M.: Neural summarization by extracting sentences and words, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 484-494 (2016)
- [Chopra 16] Chopra, S., Auli, M., and Rush, A.M.: Abstractive sentence summarization with attentive recurrent neural networks, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 93-98 (2016)
- [Clark 18] Clark, C. and Gardner, M.: Simple and effective multi-paragraph reading comprehension, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 845-855 (2018)

- [Clarke 08] Clarke, J. and Lapata, M.: Global inference for sentence compression: An integer linear programming approach, *Journal of Artificial Intelligence Research*, Vol. 31, pp. 399-429 (2008)
- [Coates-Stephens 91] Coates-Stephens, S.: Automatic acquisition of proper noun meanings, in *Proceedings of the 6th International Symposium on Methodologies for Intelligent Systems*, pp. 306-315 (1991)
- [Coster 11] Coster, W. and Kauchak, D.: Learning to simplify sentences using Wikipedia, in *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pp. 1-9 (2011)
- [Degorski 08] Degorski, L., Marcinczuk, M., and Przepiorkowski, A.: Definition extraction using a sequential combination of baseline grammars and machine learning classifiers, in *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pp. 837-841 (2008)
- [Deschacht 09] Deschacht, K. and Moens, M.: The latent words language model, in *Proceedings of the 18th Annual Belgian-Dutch Conference on Machine Learning*, pp. 1-2 (2009)
- [Devlin 98] Devlin, S. and Tait, J.: The use of a psycholinguistic database in the simplification of text for aphasic readers, *Linguistic Databases*, pp. 161-173 (1998)
- [Devlin 18] Devlin, J., Chang, M.W., Lee, K., and Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding, *arXiv:1810.04805*, pp. 1-14 (2018)
- [Dhillon 11] Dhillon, P.S., Foster, D., and Ungar, L.: Multi-view learning of word embeddings via CCA, in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pp. 199-207 (2011)
- [Do 11] Do, Q.X., Chan, Y.S., and Roth, D.: Minimally supervised event causality identification, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 294-303, 2011.
- [Doddington 02] Doddington, G.: Automatic evaluation of machine translation quality using n-gram co-occurrence statistics, in *Proceedings of the 2nd International Conference on Human Language Technology Research*. pp. 138-145 (2002)
- [Edmundson 69] Edmundson, H.P.: New methods in automatic extracting, *Journal of the ACM*, Vol. 16, No. 2, pp. 264-285 (1969)

- [Egashira 11] 江頭勇佑, 柴田知秀, 黒橋禎夫: Web から獲得した知識に基づく雑談対話システム, 平成 23 年度情報処理学会関西支部支部大会講演論文集, pp. 1-5 (2011)
- [江口 11] 江口萌, 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治: モダリティ、真偽情報、価値情報を統合した拡張モダリティ解析, 言語処理学会第 16 回年次大会論文集, pp. 852-855 (2010)
- [江原 02] 江原暉将, 和田裕二, 福島孝博: 聴覚障害者向け字幕放送における自動要約, 情報処理, テキスト自動要約特集, Vol.43, No.12, pp.1305-1309, 2002.
- [Erdelez 05] Erdelez, S.: Information encountering, in [3], pp. 179-184 (2005)
- [Erkan 04] Erkan, G. and Radev, D.R.: LexRank: Graph-based lexical centrality as salience in text summarization, *Journal of Artificial Intelligence Research*, pp. 457-479 (2004)
- [Espinosa-Anke 13] Espinosa-Anke, L.: Towards definition extraction using conditional random fields, in *Proceedings of the Student Research Workshop associated with RANLP 2013*, pp. 63-70 (2013)
- [Espinosa-Anke 14] Espinosa-Anke, L. and Saggion, H.: Applying dependency relations to definition extraction, *Natural Language Processing and Information Systems Lecture Notes in Computer Science*, Vol. 8455, pp. 63-74 (2014)
- [Fatma 17] Fatma, N., Chinnakotla, M.K., and Shrivastava, M.: The unusual suspects: Deep learning based mining of interesting entity trivia from knowledge graphs, in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 1107-1113 (2017)
- [Filatova 04] Filatova, E. and Hatzivassiloglou, V.: A formal model for information selection in multi-sentence text extraction, in *Proceedings of the 20th International Conference on Computational Linguistics*, pp. 397-403 (2004)
- [Filippova 15] Filippova, K., Alfonseca, E., Colmenares, C.A., Kaiser, L., and Vinyals, O.: Sentence compression by deletion with LSTMs, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 360-368 (2015)
- [Fisher 05] Fisher, K.E., Erdelez, S., and EmKechinie, L.E.F.: *Theories of Information Behavior*, Information Today, Inc. (2005)

- [藤江 03] 藤江真也, 江尻康, 菊池英明, 小林哲則: パラ言語の理解能力を有する対話ロボット, 情報処理学会研究報告音声言語情報処理 (SLP), Vol. 2003, No. 104(2003-SLP-048), pp. 13-20 (2003)
- [藤井 02] 藤井敦, 石川徹也: World Wide Web を用いた事典知識情報の抽出と組織化, 電子情報通信学会論文誌 D, Vol. J85-D2, No. 2, pp. 300-307 (2002)
- [福岡 17a] 福岡維新, 岩田和彦, 小林哲則: 発話系列を扱う会話音声合成, 電子情報通信学会技術研究報告, 音声研究会, Vol. 116 No. 414, pp. 59-64 (2017)
- [Fukuoka 17b] Fukuoka, I., Iwata, K., and Kobayashi, T.: Prosody control of utterance sequence for information delivering, in *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, pp. 774-778 (2017)
- [Garcia 97] Garcia, D.: COATIS, an NLP system to locate expressions of actions connected by causality links, in *Proceedings of the 10th European Workshop on Knowledge Acquisition, Modeling and Management*, pp. 347-352 (1997)
- [Gillick 08] Gillick, D., Favre, B., and Hakkani-Tur, D.: The ICSI summarization system at TAC 2008, *Text Analysis Conference* (2008)
- [Glavas 15] Glavas, G. and Stajner, S.: Simplifying lexical simplification: Do we need simplified Corpora?, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 63-68 (2015)
- [Goldensohn 03] Goldensohn, S.B., McKeown, K., and Schlaikjer, A.: Defscriber: a hybrid system for definitional QA, in *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 462-462 (2003)
- [Goldstein 00] Goldstein, J., Mittal, V., Carbonell, J., and Kantrowitz, M.: Multi-Document Summarization by Sentence Extraction, in *Proceedings of the ANLP/NAACL 2000 Workshop: Automatic Summarization*, pp. 40-48 (2000)
- [後藤 14] 後藤功雄, 熊野正, 田中英輝: 一般のニュースからやさしい日本語ニュースへの書き換えの分析, 言語処理学会第 20 回年次大会発表論文集, pp. 15-18 (2014)
- [後藤 16] 後藤功雄, 田中英輝: 日本語連体修飾節の抽出による文分割の可否判断に適した基準の提案, 言語処理学会 第 22 回年次大会 発表論文集, pp. 1057-1060 (2016)

- [Granroth-Wilding 16] Granroth-Wilding, M. and Clark, S.: What happens next? event prediction using a compositional neural network model, in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 2727-2733 (2016)
- [Graves 14] Graves, A., Wayne, G., and Danihelka, I.: Neural turing machines, *arXiv:1410.5401*, pp. 1-26 (2014)
- [Graves 16] Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Barwinska, A.G., Colmenarejo, S.G., Grefenstette, E., Ramalho, T., Agapiou, J., Badia, A.P., Hermann, K.M., Zwols, Y., Ostrovski, G., Cain, A., King, H., Summerfield, C., Blunsom, P., Kavukcuoglu, K., and Hassabis, D.: Hybrid computing using a neural network with dynamic external memory, *Nature*, Vol. 538, pp. 471-476 (2016)
- [Gulcehre 16] Gulcehre, C., Ahn, S., Nallapati, R., Zhou, B., and Bengio, Y.: Pointing the unknown words, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 140-149 (2016)
- [Guo 18] Guo, L., Wang, L., Dang, J., Zhang, L., Guan, H., and Li, X.: Speech emotion recognition by combining amplitude and phase information using convolutional neural network, in *Proceedings of the 19th Annual Conference of the International Speech Communication Association*, pp. 1611-1615 (2018)
- [幅田 01] 幅田隆, 奥村学: 不要個所削除による講演音声の要約, 言語処理学会第7回年次大会発表論文集, pp. 289-292 (2001)
- [埴 17] 埴一晃, 佐々木彬, 岡崎直観, 乾健太郎: クラウドソーシングによる関係知識のアノテーション, 言語処理学会第23回年次大会, pp. 406-409 (2017)
- [Harabagiu 02] Harabagiu, S., Moldovan, D., and Picone, J.: Open-domain voice-activated question answering, in *Proceedings of the 19th international conference on Computational linguistics*, pp. 1-7 (2002)
- [原田 01] 原田実, 水野高宏: EDR を用いた日本語意味解析システム SAGE, 人工知能学会論文誌, Vol. 16, No. 1, pp. 85-93 (2001)
- [原田 02] 原田実, 田淵和幸, 大野博之: 日本語意味解析システム SAGE の高速化・高精度化とコーパスによる精度評価, 情報処理学会論文誌, Vol. 43, No. 9, pp. 2894-2902 (2002)

- [Harada 07] Harada, M., Kato, Y., Takehara, K., Kawamata, M., Sugimura, K., and Kawaguchi, J.: QA system Metis based on semantic graph matching, in *Proceedings of the 6th International Conference on NII Test Collection for IR System*, pp. 448-459 (2007)
- [原田 11] 原田実, 西岡晋太郎: 深層格とオントロジーを用いた回答抽出による質問応答システム Metis の回答精度向上, 情報処理学会第 73 回全国大会, pp. 49-50 (2011)
- [Hasan 16] Hasan, A.M. and Zakaria, L.Q.: Question classification using support vector machine and pattern matching, in *Journal of Theoretical and Applied Information Technology*, Vol. 87, No. 2, pp. 259-265 (2016)
- [橋本 04] 橋本泰一, 乾孝司, 村上浩司: 拡張固有表現タグ付きコーパスの構築, 情報処理学会論文誌, Vol. 45, No. 3, pp. 934-941 (2004)
- [Hashimoto 09] Hashimoto, C., Torisawa, K., Kuroda, K., Murata, M., and Kazama, J.: Large-scale verb entailment acquisition from the web, in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 1172-1181 (2009)
- [橋本 10] 橋本力, 鳥澤健太郎, 黒橋禎夫, ステイン・デ・サーガ, 風間淳一, 藤田篤: 動詞含意関係データベースの自動拡張, 言語処理学会第 16 回年次大会発表論文集, pp. 940-943 (2010)
- [橋本 11a] 橋本力, 鳥澤健太郎, 黒田航, デサーガスティン, 村田真樹, 風間淳一: WWW からの大規模動詞含意知識の獲得, 情報処理学会論文誌, Vol.52, No.1, pp.293-307, 2011.
- [橋本 11b] 橋本力, 鳥澤健太郎, ステイン・デ・サーガ, 風間淳一, 黒橋禎夫: Web 上の定義文からの言い換え知識獲得, 言語処理学会第 17 回年次大会論文集, pp.748-751 (2011)
- [Hashimoto 11c] Hashimoto, C., Torisawa, K., Saeger, S.D., Kazama, J., and Kurohashi, S.: Extracting paraphrases from definition sentences on the web, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 1087-1097 (2011)
- [橋本 12a] 橋本力, 鳥澤健太郎, Stijn De Saeger, 呉鍾勲, 風間淳一: もう一つの意味的極性「活性／不活性」と知識獲得への応用, 言語処理学会第 18 回年次大会論文集, pp. 93-96 (2012)
- [Hashimoto 12b] Hashimoto, C., Torisawa, K., Saeger, S.D., Oh, J.H., and Kazama, J.: Excitatory or Inhibitory: A new semantic orientation extracts contradiction and causality from the web, in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 619-630 (2012)

- [Hashimoto 12c] Hashimoto, C., Torisawa, K., Saeger, S.D., Oh, J.H., and Kazama, J.: Supplementary material for the paper: Excitatory or Inhibitory: A new semantic orientation extracts contradiction and causality from the web, pp. 1-10 (2012)
- [橋本 12d] 橋本力: NICT 発の基盤的言語資源 - 言い換え処理向け言語資源を中心に -, *Japio year book*, pp. 212-217 (2012)
- [橋本 12e] 橋本力, 呉鍾勲, 佐野大樹, 川田拓也: 5-5 基盤的言語資源, 知識創成コミュニケーション特集, 情報通信研究機構季報, Vol. 58, Nos. 3/4, pp. 113-136 (2012)
- [Hashimoto 14] Hashimoto, C., Torisawa, K., Kloetzer, J., Sano, M., Varga, I., Oh, J.H. and Kidawara, Y.: Toward future scenario generation : Extracting event causality exploiting semantic relation, context, and association features, in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 987-997 (2014)
- [Hashimoto 14] Hashimoto, C., Torisawa, K., Kloetzer, J., Sano, M., Varga, I., Oh, J.H., and Kidawara, Y.: Supplementary notes for paper: Toward future scenario generation : Extracting event causality exploiting semantic relation, context, and association features, pp. 1-26 (2014)
- [Hashimoto 15] Hashimoto, C., Torisawa, K., Kloetzer, J., and Oh, J.H.: Generating event causality hypotheses through semantic relations, in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp. 2396-2403 (2015)
- [畠 87] 畠弘巳: 話しことばの特徴 - 冗長性をめぐって -, 国文学解釈と鑑賞, Vol. 52, No. 7, pp. 22-34 (1987)
- [林 07a] 林由紀子, 松原茂樹: 聞きやすい読み上げ音声出力のためのテキスト変換の検討, 情報処理学会第 69 回全国大会講演論文集, Vol. 2, pp. 581-582 (2007)
- [林 07b] 林由紀子, 松原茂樹: 自然な読み上げ音声出力のための書き言葉から話し言葉へのテキスト変換, 情報処理学会研究報告, 音声言語情報処理, Vol. 47, pp. 49-54 (2007)
- [林 08] 林由紀子, 松原茂樹: ニュース記事の自然な音声出力のためのテキスト変換, 言語処理学会第 14 回年次大会発表論文集, pp. 790-793 (2008)
- [林 14] 林佑樹, 大佛駿介, 中野有紀子: 協調学習における韻律特徴を用いた発話タグ推定モデル, 教育システム情報学会第 39 回全国大会, pp. 441-442 (2014)

- [Hegde 07] Hegde, R.M., Murthy, H.A., and Gadde, V.R.R.: Significance of the modified group delay feature in speech recognition, *IEEE Transactions on Audio Speech and Language Processing*, Vol. 15, No. 1, pp. 190-202 (2006)
- [Higashinaka 08a] Higashinaka, R. and Isozaki, H.: Automatically acquiring causal expression patterns from relation-annotated corpora to improve question answering for why-questions, *ACM Transactions on Asian Language Information Processing, Special Issue on NTCIR-6*, Vol. 7, No. 6, pp. 1-27 (2008)
- [Higashinaka 08b] Higashinaka, R. and Isozaki, H.: Corpus-based question answering for why-questions, in *Proceedings of the 3rd International Joint Conference on Natural Language Processing*, pp. 418-425 (2008)
- [Higashinaka 12] Higashinaka, R., Sadamitsu, K., Makino, T., Matsuo, Y., and Saito, K.: Creating an extended named entity dictionary from wikipedia, in *Proceedings of the 24th International Conference on Computational Linguistics*, pp. 1163-1178 (2012)
- [東中 13a] 東中竜一郎, 貞光九月, 齋藤邦子, 小林のぞみ: 幅広い質問にピンポイントで答える質問応答技術, NTT 技術ジャーナル, Vol. 20, No. 4, pp. 25-28 (2013)
- [東中 13b] 東中竜一郎, 貞光九月, 内田渉, 吉村健: シャベってコンシェルにおける質問応答技術, NTT 技術ジャーナル, Vol. 25, No.2 (2013)
- [東山 08] 東山昌彦, 乾健太郎, 松本裕治, 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第 14 回年次大会論文集, pp. 584-587 (2008)
- [Hirano 10] Hirano, T., Asano, H., Matsuo, Y., and Kikui, G.: Recognizing relation expression between named entities based on inherent and context-dependent features of relational words, in *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 409-417 (2010)
- [Hirao 02] Hirao, T., Isozaki, H., Maeda, E., and Matsumoto, Y.: Extracting important sentences with support vector machines, in *Proceedings of the 19th international conference on Computational linguistics*, pp. 1-7 (2002)
- [Hirao 13] Hirao, T., Yoshida, Y., Nishino, M., Yasuda, N., and Nagata, M.: Single-document summarization as a tree knapsack problem, in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1515-1520 (2013)

- [弘前大学] 「やさしい日本語」, 弘前大学人文学部社会言語学研究室, <http://human.cc.hirosaki-u.ac.jp/kokugo/EJ3mokuji.htm>
- [広瀬 10] 広瀬和生: 現代落語の基礎知識, 集英社 (2010)
- [Hirschberg 96] Hirschberg, J. and Nakatani, C.H.: A prosodic analysis of discourse segments in direction-giving monologues, in *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*, pp. 286-293 (1996)
- [Hoard 92] Hoard, J.E., Wojcik, R., and Holzhauser, K.: An automated grammar and style checker for writers of simplified English, *Computers and Writing*, Springer Netherlands, pp. 278-296 (1992)
- [北条 16a] 北条伸克, 井島勇祐, 杉山弘晃: 対話行為情報を表現可能な音声合成の検討, 人工知能学会 第 30 回全国大会 論文集, 204-OS-23a-4, pp. 1-4 (2016)
- [Hojo 16b] Hojo, N., Ijima, Y., and Mizuno, H.: An investigation of DNN-based speech synthesis using speaker codes, in *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, pp. 2278-2282 (2016)
- [Hojo 18] Hojo, N., Ijima, Y., and Mizuno, H.: DNN-based speech synthesis using speaker codes, *IEICE Transactions on Information and Systems*, Vol. E101D, No. 2, pp. 462-472 (2018)
- [堀 01] 堀智織, 古井貞熙: 講演音声の自動要約の試み, 話し言葉の科学と工学ワークショップ 講演予稿集, pp. 165-171 (2001)
- [Hori 03] Hori, C., Hori, T., Isozaki, H., Maeda, E., Katagiri, S., and Furui, S.: Deriving disambiguous queries in a spoken interactive ODQA system, in *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 624-627 (2003)
- [星野 18] 星野綾子, 細見格: 句構造解析とクラスタリングを用いた会話履歴の要約, 人工知能学会第 32 回全国大会論文集, 2K1-02, pp. 1-3 (2018)
- [Hovy 97] Hovy, E. and Lin, C.Y.: Automated text summarization in SUMMARIST, in *Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*, pp. 18-24 (1997)

- [Hromkovic 03] Hromkovic, J.: Algorithmics for hard problems, Springer (2003)
- [Hu 18] Hu, M., Wei, F., Peng, Y., Huang, Z., Yang, N., and Li, D.: Read + Verify: Machine reading comprehension with unanswerable questions, *arXiv:1808.05759*, pp. 1-9 (2018)
- [飯田 10] 飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕治: 述語項構造と照応関係のアノテーション: NAIST テキストコーパス構築の経験から, 自然言語処理, Vol. 17, No. 2, pp. 25-50 (2010)
- [Iida 11] Iida, R. and Poesio, M.: A cross-lingual ILP solution to zero anaphora resolution, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 804-813 (2011)
- [今田 14] 今田水穂: 拡張固有表現階層から SUMO への対応表, 第 6 回コーパス日本語学ワークショップ予稿集, pp. 183-192 (2014)
- [稲葉 14] 稲葉通将, 神園彩香, 高橋健一: Twitter を用いた非タスク指向型対話システムのための発話候補文獲得, 人工知能学会論文誌, Vol. 29, No. 1, pp. 21-31 (2014)
- [Inoue 17] Inoue, K., Hara, S., Abe, M., Hojo, N., and Ijima, Y.: An investigation to transplant emotional expressions in DNN-based TTS synthesis, in *Proceedings of the 9th Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1253-1258 (2017)
- [乾 04] 乾考司, 乾健太郎, 松本裕治: 接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得, 情報処理学会論文誌, Vol. 45, No. 3, pp. 919-933 (2004)
- [Isonuma 17] Isonuma, M., Fujino, T., Mori, J., Matsuo, Y., and Sakata, I.: Extractive summarization using multi-task learning with document classification, in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2101-2110 (2017)
- [磯崎 08] 磯崎秀樹, 東中竜一郎: パターンマイニングを用いて「なぜ」に答えるシステム, 言語処理学会第 14 回年次大会発表論文集, pp. 1025-1028 (2008)
- [磯崎 09] 磯崎秀樹, 東中竜一郎: Why 型質問応答システム NAZEQA Web 版, 情報処理学会関西支部支部大会講演論文集, pp. 143-146 (2009)
- [岩畑 10] 岩畑貴弘: 「情報のなわ張り理論」再考, 神奈川大学人文研究, 2010.

- [岩倉 11] 岩倉友哉, 高村大也, 奥村学: チャンクの分解・結合に基づく拡張固有表現抽出手法, 言語処理学会第 17 回年次大会発表論文集, pp. 786-789 (2011)
- [岩田 94] 岩田和彦: 文章朗読音声における韻律的特徴の分析, 日本音響学会研究発表会講演論文集, pp. 285-286 (1994)
- [Iwata 11] Iwata, K. and Kobayashi, T.: Conversational speech synthesis system with communication situation dependent HMMs, in *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*, pp. 113-123 (2011)
- [岩田 12] 岩田和彦, 小林哲則: 終助詞とその音調とによって聞き手に伝わる発話意図の分析, 電子情報通信学会技術研究報告, Vol. 112, No. 281, pp. 31-36 (2012)
- [Izumi 14a] Izumi, T., Shibata, T., Asano, H., Matsuo, Y., and Kurohashi, S.: Constructing a corpus of Japanese predicate phrases for synonym/antonym relations, in *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pp. 1394-1400 (2014)
- [泉 14b] 泉朋子, 柴田知秀, 浅野久子, 松尾義博, 黒橋禎夫: 述部意味関係コーパスの構築, 言語処理学会第 20 回年次大会発表論文集, pp. 690-693 (2014)
- [Jean 14] Jean, S., Cho, K., Memisevic, R., Bengio, Y.: On using very large target vocabulary for neural machine translation, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 1-10 (2014)
- [Jelinek 69] Jelinek, F.: Fast sequential decoding algorithm using a stack, *IBM Journal of Research and Development*, Vol. 13, pp. 675-685 (1969)
- [Joho 99] Joho, H. and Joho, H.: Retrieving descriptive phrases from large amounts of free text, in *Proceedings of the ninth international conference on Information and knowledge management*, pp. 180-186 (2000)
- [Joshi 17] Joshi, M., Choi, E., Weld, D.S., and Zettlemoyer, L.: TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1601-1611 (2017)

- [Joskowicz 89] Joskowicz, L., Ksiezzyk, T., and Grishman, R.: Deep domain models for discourse analysis, in *Proceedings of the Annual AI Systems in Government Conference*, pp. 195-200 (1989)
- [Joulin 17] Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T.: Bag of tricks for efficient text classification, in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Vol. 2, pp. 427-431 (2017)
- [カーネマン 14] ダニエル・カーネマン: ファスト&スロー あなたの意思はどのように決まるか?, ハヤカワ・ノンフィクション文庫, 早川書房 (2014)
- [海木 96] 海木延佳, 匂坂芳典: 局所的な句構造によるポーズ挿入規則化の検討, 信学論 (D-II), Vol. J79-D-II, No. 9, pp. 1455-1463 (1996)
- [鍛冶 04] 鍛冶伸裕, 岡本雅史, 黒橋禎夫: WWW を用いた書き言葉特有語彙から話し言葉語彙への用言の言い換え, 自然言語処理学会論文誌, Vol. 11, No. 5, pp. 19-37 (2004)
- [Kaji 06] Kaji, N. and Kitsuregawa, M.: Automatic construction of polarity-tagged corpus from HTML documents, in *Proceedings of the COLING/ACL on Main conference poster sessions*, pp. 452-459 (2006)
- [Kaji 07] Kaji, N. and Kitsuregawa, M.: Building lexicon for sentiment analysis from massive collection of HTML documents, in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 1075-1083 (2007)
- [梶原 17] 梶原智之, 小町守: Simple PPDB: Japanese, 言語処理学会第 23 回年次大会発表論文集, pp. 529-532 (2017)
- [神尾 90] 神尾昭雄: 情報のなわ張り理論, 大修館書店, 1990.
- [神尾 02] 神尾昭雄: 続・情報のなわ張り理論, 大修館書店, 2002.
- [鴨井 11] 鴨井一人, 山本知仁, 三宅美博: 文章発話におけるポーズ長とその前後の発話長の作用関係, ヒューマンインタフェースシンポジウム 2011 論文集, pp. 357-364 (2011)
- [金山 11a] 金山博, 武田浩一: Watson: クイズ番組に挑戦する質問応答システム, 情報処理, Vol. 52, No. 7 (2011)

- [金山 11b] 金山博, 武田浩一: 質問応答システム Watson が示す未来, *PROVISION*, No. 70 (2011)
- [Kauchak 13] Kauchak, D.: Improving text simplification language modeling using unsimplified text data, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 1537-1546 (2013)
- [川口 09] 川口純一, 青木洋, 松田源立, 原田実: 意味解析システム SAGE の精度向上, pp. 77-78 (2009)
- [河原 04] 河原達也: 話し言葉による音声対話システム, *IPSJ Magazine*, Vol. 45, No. 10, pp. 1027-1031 (2004)
- [Kawahara 06] Kawahara, H.: STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds, *Acoustical Science and Technology*, Vol. 27, No. 6, pp. 349-353 (2006)
- [Kawahara 06b] Kawahara, D. and Kurohashi, S.: A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis, in *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pp. 176-183 (2006)
- [河原 07] 河原大輔, 黒橋禎夫: 自動構築した大規模格フレームに基づく構文・格解析の統合的
確率モデル, *自然言語処理*, Vol. 14, No. 4, pp. 67-81 (2007)
- [河原 08] 河原達也, 川嶋宏彰, 平山高嗣, 松山隆司: 対話を通じてユーザの意図・興味を探り情
報検索・提示する情報コンシェルジェ, *情報処理*, Vol. 49, No. 8, pp. 1-6 (2008)
- [川村 12] 川村よし子: 日本語読解学習システム『リーディング・チュウ太』の取り組み, ウェ
ブマガジン「留学交流」, Vol. 19, pp. 1-11 (2012)
- [風間 08a] 風間淳一, 鳥澤健太郎: Web 上の資源から構築した複数の固有表現辞書を用いた日
本語固有表現認識, *言語処理学会第 14 回年次大会発表論文集*, pp. 813-816 (2008)
- [Kazama 08b] Kazama, J. and Torisawa, K.: Inducing gazetteers for named entity recognition
by large-scale clustering of dependency relations, in *Proceedings of the 46th Annual Meeting
of the Association for Computational Linguistics: Human Language Technologies*, pp. 407-
415 (2008)

- [風間 09] 風間淳一, デサーガ・ステイン, 鳥澤健太郎, 村田真樹: 係り受けの確率的クラスタリングを用いた大規模類似語リストの作成, 言語処理学会第 15 回年次大会発表論文集, pp. 84-87 (2009)
- [芥子 17] 芥子育雄, 鈴木優, 吉野幸一郎, グラムニュービッツ, 大原一人, 向井理朗, 中村哲: 単語意味ベクトル辞書を用いた Twitter からの日本語評判情報抽出, 電子情報通信学会論文誌, Vol. J100-D, No. 4, pp. 530-543 (2017)
- [Khoo 00] Khoo, C.S.G., Chan S., and Niu, Y.: Extracting causal knowledge from a medical database using graphical patterns, in *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pp. 336-343 (2000)
- [Khuller 99] Khuller, S., Moss, A., and Naor, J. S.: The budgeted maximum coverage problem, *Information Processing Letters*, Vol. 70, No. 1, pp. 39-45 (1999)
- [Khuller 06] Khuller, S., Raschid, L., and Wu, Y.: LP randomized rounding for maximum coverage problem and minimum set cover with threshold problem, *Technical Report, CS-TR-4805, The University of Maryland*, pp. 1-12 (2006)
- [木俵 12] 木俵豊: 情報分析システム WISDOM の開発, 情報通信研究機構季報, Vol.5, Nos.3/4, 2012.
- [Kikuchi 14] Kikuchi, Y., Hirao, T. Takamura, H., Okumura, M., Nagata, M.: Single document summarization based on nested tree structure, in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 315-320 (2014)
- [菊池 15] 菊池悠太, 平尾努, 高村大也, 奥村学, 永田昌明: 入れ子依存木の刈り込みによる単一文書要約手法, 自然言語処理, Vol. 22, No. 3, pp. 197-217 (2015)
- [Kikuchi 16] Kikuchi, Y., Neubig, G., Sasano, R., Takamura, H., and Okumura, M.: Controlling output length in neural encoder-decoders, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1328-1338 (2016)
- [Kim 14] Kim, Y.: Convolutional neural networks for sentence classification, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746-1751 (2014)

- [Kincaid 75] Kincaid, J., Fishburne, R., Rogers, R., and Chissom, B.: Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for navy enlisted personnel, *Research Branch report*, pp. 1-49 (1975)
- [Kiros 15] Kiros, R., Zhu., Y., Salakhutdinov, R., Zemel, R.S., Torralba, A., Urtasun, R., and Fidler, S.: Skip-thought vectors, in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, pp. 3294-3302 (2015)
- [北島 12] 北島理沙, 小林一郎: 潜在的意味を考慮したグラフに基づく複数文書要約, ARG 第 1 回 Web インテリジェンスとインタラクション研究会, ARG-WI2-1-21, pp. 1-6 (2012)
- [清野 18] 清野舜, 高瀬翔, 鈴木潤, 岡崎直観, 乾健太郎, 永田昌明: ニューラルヘッドライン生成における誤生成問題の改善, 言語処理学会第 24 回年次大会発表論文集, pp. 1-4 (2018)
- [Klerke 16] Klerke, S., Goldberg, Y., and Sogaard, A.: Improving sentence compression by learning to predict gaze, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1528-1533 (2016)
- [Kloetzer 13a] Kloetzer, J., Saeger, S.D., Torisawa, K., Sano, M., Hashimoto, C., and Gotoh, J.: Large-scale acquisition of entailment pattern pairs, in *Information Processing Society of Japan (IPSJ) Kansai-Branch Convention*, pp. 1-6 (2013)
- [Kloetzer 13b] Kloetzer, J., Saeger, S.D., Torisawa, K., Hashimoto, C., Oh, J.H., Sano, M., and Ohtake, K.: Two-stage method for large-scale acquisition of contradiction pattern pairs using entailment, in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 693-703 (2013)
- [Kloetzer 14] Julien Kloetzer, 鳥澤健太郎, Stijn De Saeger, 佐野大樹, 橋本力, 呉鍾勲, 大竹清敬: 含意関係を用いたパターン間矛盾関係獲得, 言語処理学会第 20 回年次大会発表論文集, pp. 344-347 (2014)
- [Knight 02] Knight, K. and Marcu, D.: Summarization beyond sentence extraction: A probabilistic approach to sentence compression, *Journal of Artificial Intelligence*, Vol. 139, pp. 91-107 (2002)
- [小林 05] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol. 12, No. 3, pp. 203-222 (2005)

- [小林 14] 小林哲則, 岩田和彦: 会話向け音声合成システム, 電子情報通信学会技術研究報告, Vol. 114, No. 303, pp. 19-24 (2014)
- [Koehn 07] Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E.: Moses: Open source toolkit for statistical machine translation, in *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pp. 177-180 (2007)
- [Koiso 97] Koiso, H., Shimojima, A., and Katagiri, Y.: Informational potentials of dynamic speech rate in dialogue, in *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, pp. 394-399 (1997)
- [小島 10] 小島正裕, 村田真樹, 風間淳一, 黒田航, 藤田篤, 荒牧英治, 土田正明, 渡辺靖彦, 鳥澤健太郎: 機械学習と種々の素性を用いた編集距離の小さい日本語異表記対の抽出, 言語処理学会第 16 回年次大会発表論文集, pp. 928-931 (2010)
- [駒谷 01] 駒谷和範, 河原達也, 清田陽司, 黒橋禎夫, Pascale Fung: 柔軟な言語モデルとマッチングを用いた音声によるレストラン検索システム, 電子情報通信学会技術研究報告, NLC2001-78, SP2001-113 (SLP-39-30), pp. 67-72 (2001)
- [駒谷 03a] 駒谷和範, 上野晋一, 河原達也, 奥乃博: ユーザモデルを導入したバス運行情報案内システムの実験的評価, 情報処理学会研究報告, Vol. 2003, No. 75(SLP-47), pp. 59-64 (2003)
- [駒谷 03b] 駒谷和範, 鹿島博晶, 田中克明, 河原達也: 複合的言語制約に基づくキーフレーズ検出を用いた汎用的なデータベース検索音声対話プラットフォーム, 情報処理学会論文誌, Vol.44, No.5, pp.1333-1342 (2003)
- [Kontos 91] Kontos, J. and Sidiropoulou, M.: On the acquisition of causal knowledge from scientific texts with attribute grammars, in *International Journal of Applied Expert Systems*, Vol.4, No. 1, pp. 31-48 (1991)
- [久保田 08] 久保田裕章, 平塚飛将, 吉川ひかる, 松田源立, 原田実: 質問応答システム Metis の回答精度向上 - 検索フェーズの改良を中心として -, 言語処理学会第 14 回年次大会発表論文集, pp. 1017-1020 (2008)
- [Kudo 00] Kudo, T. and Matsumoto, Y.: Japanese dependency analysis based on support vector machines, in *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods*

in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics, pp. 18-25 (2000)

[Kudo 02a] Kudo, T. and Matsumoto, Y.: Japanese dependency analysis using cascaded chunking, in *Proceedings of the 6th Conference on Natural Language Learning*, pp. 1-7 (2002)

[工藤 02b] 工藤拓, 松本裕治: チャンキングの段階適用による係り受け解析, 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834-1842 (2002)

[Kudo 04] Kudo, T., Yamamoto, K., and Matsumoto, Y.: Applying conditional random fields to Japanese morphological analysis, in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 230-237 (2004)

[Kumar 14] Kumar, A.M. and Soman, K.P.: A conditional random field framework for Thai morphological analysis, in *Proceedings of the Forum for Information Retrieval Evaluation*, pp. 112-120 (2014)

[Kupiec 95] Kupiec, J., Pedersen, J.O., Chen, F.: A trainable document summarizer, in *Proceedings of the 18th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 68-73 (1995)

[栗原 18] 栗原理聡, 水本智也, 乾健太郎: Twitterによる評判分析を目的とした評価対象-評価表現データセット作成, 言語処理学会第24回年次大会発表論文集, pp. 344-347 (2018)

[黒田 10] 黒田航, 風間淳一, 村田真樹, 鳥澤健太郎: Webデータに対応できる日本語異表記対の認定基準, 言語処理学会第16回年次大会発表論文集, pp. 990-993 (2010)

[黒田 09] 黒田航, 李在鎬, 野澤元, 村田真樹, 鳥澤健太郎: 鳥式改の上位語データの手クリーニング, 言語処理学会15回年次大会発表論文集, pp. 76-79 (2009)

[Lafferty 01] Lafferty, J., McCallun, A., and Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data, in *Proceedings of the 18th International Conference on Machine Learning*, pp. 282-289 (2001)

[Lample 16] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C.: Neural architectures for named entity recognition, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 260-270 (2016)

- [Lapata 03] Lapata, M.: Probabilistic text structuring: Experiments with sentence ordering, in *Proceedings of the 41st Meeting of the Association of Computational Linguistics*, pp. 545-552 (2003)
- [Le 14] Le, Q. and Mikolov, T.: Distributed representations of sentences and documents, in *Proceedings of the 31st International Conference on Machine Learning*, pp. 1188-1196 (2014)
- [Lebret 14] Lebret, R. and Collobert, R.: Word embeddings through hellinger PCA, in *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 482-490 (2014)
- [Levy 17] Levy, O., Seo, M., Choi, E., and Zettlemoyer, L.: Zero-shot relation extraction via reading comprehension, in *Proceedings of the 21st Conference on Computational Natural Language Learning*, pp. 333-342 (2017)
- [Lin 97] Lin, C.Y. and Hovy, E.: Identifying Topics by Position, in *Proceedings of the 5th Conference on Applied Natural Language Processing*, pp. 283-290 (1997)
- [Lin 99] Lin, C.Y.: Training a selection function for extraction, in *Proceedings of the Eighteenth Annual International ACM Conference on Information and Knowledge Management*, pp. 1-8 (1999)
- [Lin 01] Lin, D. and Pantel, P.: Discovery of inference rules for question answering, *Natural Language Engineering*, Vol. 7, No. 4, pp. 343-360 (2001)
- [Lin 03] Lin, C. and Hovy, E.: Automatic evaluation of summaries using n-gram co-occurrence statistics, in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Vol. 1, pp. 71-78 (2003)
- [Lin 04] Lin, C.: ROUGE: A package for automatic evaluation of summaries, in *Proceedings of the ACL-04 Workshop: Text Summarization Branches Out*, pp. 74-81 (2004)
- [Lin 15] Lin, R., Liu, S., Yang, M., Li, M., Zhou, M., and Li, S.: Hierarchical recurrent neural network for document modeling, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 899-907 (2015)

- [Lin 17] Lin, Z., Feng, M., Santos, C.N., Yu, M., Xiang, B., Zhou, B., and Bengio, Y.: A structured self-attentive sentence embedding, in *Proceedings of the 5th International Conference on Learning Representations*, pp. 1-15 (2017)
- [Luhn 58] Luhn, H.P.: The automatic creation of literature abstracts, *IBM Journal of Research and Development*, Vol. 2, No. 2, pp. 159-165 (1958)
- [Luo 18] Luo, D., Zou, Y., and Huang, D.: Investigation on joint representation learning for robust feature extraction in speech emotion recognition, in *Proceedings of the 19th Annual Conference of the International Speech Communication Association*, pp. 152-156 (2018)
- [Luong 17] Luong, H.T., Takaki, S., Henter, G.E., and Yamagishi, J.: Adapting and controlling DNN-based speech synthesis using input codes, in *Proceedings of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1905-1909 (2017)
- [前澤 04] 前澤敏之, 面来道彦, 上野雅和, 韓東力, 原田実: 意味解析システム SAGE の高精度化と概念グラフへの変換, 情報処理学会第 66 回全国大会, pp. 177-178 (2004)
- [Malisz 17] Malisz, Z., Berthelsen, H., Beskow, J., and Gustafson, J.: Controlling prominence realisation in parametric DNN-based speech synthesis, in *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, pp. 1079-1083 (2017)
- [Mani 97] Mani, I. and Bloedorn, E.: Multi-document summarization by graph search and matching, in *Proceedings of the 14th National Conference on Artificial Intelligence*, pp. 622-628 (1997)
- [Mani 98] Mani, I., Bloedorn, E., and Gates, B.: Using cohesion and coherence models for text summarization, in *Proceedings of the Intelligent Text Summarization Symposium, AAAI Press, Technical Report SS-98-06*, pp. 60-67 (1998)
- [Mann 88] Mann, W.C. and Thompson, S.A.: Rhetorical structure theory: Toward a functional theory of text organization, *Text-Interdisciplinary Journal for the Study of Discourse*, Vol. 8, No. 3, pp. 243-281 (1988)
- [Marcu 97] Marcu, D.: From discourse structures to text summaries, in *Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*, pp. 82-88 (1997)
- [Marcu 98] Marcu, D.: Improving summarization through rhetorical parsing tuning, in *Proceedings of the COLING-ACL, Workshop on Very Large Corpora*, pp. 10-16 (1998)

- [Maruyama 18] Maruyama, T. and Yamamoto, K.: Simplified corpus with core vocabulary, in *Proceedings of the 11th International Conference on Language Resources and Evaluation*, pp. 1153-1160 (2018)
- [Matsuyama 14] Matsuyama, Y., Papangelis, A., Zhao, R., and Cassell, J.: 2 者会話におけるラポール形成・維持・崩壊の計算モデル, 人工知能学会研究会資料, SIG-SLUD-B402-03, pp. 13-18 (2014)
- [Matsuyoshi 10] Matsuyoshi, S., Eguchi, M., Sao, C., Murakami, K., Inui, K., and Matsumoto, Y.: Annotating event mentions in text with modality, focus, and source information. in *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, pp. 1456-1463 (2010)
- [松吉 11] 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治: 拡張モダリティタグ付与コーパスの設計と構築, 言語処理学会第 17 回年次大会論文集, pp. 147-150 (2011)
- [McDonald 07] McDonald, R.: A study of global inference algorithms in multi-document summarization, in *Proceedings of the 29th European Conference on Information Retrieval*, pp. 557-564 (2007)
- [McLaughlin 69] McLaughlin, G.H.: Smog grading: A new readability formula, *Journal of reading*, Vol. 12, No. 8, pp. 639-646 (1969)
- [目黒 14] 目黒豊美, 杉山弘晃, 東中竜一郎, 南泰浩: ルールベース発話生成と統計的発話生成の融合に基づく対話システムの構築, 人工知能学会第 28 回全国大会論文集, 2M5-OS-20b-2, pp. 1-4 (2014)
- [Mihalcea 04] Mihalcea, R. and Tarau, P.: TextRank: Bringing order into texts, in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 401-411 (2004)
- [Miike 94] Miike, S., Itoh, E., Ono, K., and Sumita, K.: A full-text retrieval system with a dynamic abstract generation function, in *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 152-161 (1994)

- [Mikolov 13a] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient estimation of word representations in vector space, in *Proceedings of the 1st International Conference on Learning Representations*, pp. 1-12 (2013)
- [Mikolov 13b] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J.: Distributed representations of words and phrases and their compositionality, in *Proceedings of the 26th International Conference on Neural Information Processing Systems*, pp.3111-3119 (2013)
- [Mikolov 13c] Mikolov, T., Yih, W., and Zweig, G.: Linguistic regularities in continuous space word representations, in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746-751 (2013)
- [Miliaraki 04] Miliaraki, S. and Androutsopoulos, I.: Learning to identify single-snippet answers to definition questions, in *Proceedings of the 20th International Conference on Computational Linguistics*, pp. 1360-1366 (2004)
- [Minjoon 17] Minjoon, S., Aniruddha, K., and Ali, F.; Hajishirzi, Hannaneh: Bidirectional attention flow for machine comprehension, in *Proceedings of the 5th International Conference on Learning Representations*, pp. 1-13 (2017)
- [美野 10] 美野秀弥, 田中英輝: 国語辞典を使った放送ニュースの名詞の平易化, 言語処理学会第16回年次大会発表論文集, pp. 760-763 (2010)
- [美野 11] 美野秀弥, 田中英輝: 放送ニュースの動詞連用形名詞の平易化, 言語処理学会第17回年次大会発表論文集, pp. 744-747 (2011)
- [美野 12] 美野秀弥, 田中英輝. ニュース原稿のやさしい日本語ニュースへの書き換え支援ツール - 日本在住外国人のために -, 映像情報メディア学会年次大会, No. 18-6, pp. 1-2 (2012)
- [Misu 07] 翠輝久, 河原達也, 正司哲朗, 美濃導彦: 質問応答・情報推薦機能を備えた音声による情報案内システム, 情処理学論, Vol.48, No.12, pp.3602-3611, 2007.
- [Mitra 97] Mitra, M., Singhal, A., and Buckley, C.: Automatic text summarization by paragraph extraction, in *Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*, pp. 39-46 (1997)
- [水野 07a] 水野淳太, 秋葉友良: 任意の回答を対象とする質問応答のための実世界質問の分析と回答タイプ判定法の検討, 言語処理学会第13回年次大会発表論文集, pp. 1002-1005 (2007)

- [Mizuno 07b] Mizuno, J., Akiba, T., Fujii, A., and Itou, K.: Non-factoid question answering experiments at NTCIR-6: Towards answer type detection for real world questions, in *Proceedings of the 6th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access*, pp. 487-492 (2007)
- [水野 09] 水野淳太, 乾健太郎, 松本裕治: ウェブニュースを利用した雑談対話システム, 言語・音声理解と対話処理研究会, Vol. 55, pp. 1-6 (2009)
- [Mizuno 16] Mizuno, J., Tanaka, M., Ohtake, K., Oh, J.H., Kloetzer, J., Hashimoto, C., and Torisawa, K.: WISDOM X, DISAANA and D-SUMM: Large-scale NLP systems for analyzing textual big data, in *Proceedings of the 26th International Conference on Computational Linguistics: System Demonstrations*, pp. 263-267 (2016)
- [水野 17] 水野淳太, 田仲正弘, 大竹清敬, 呉鍾勲, Julien Kloetzer, 橋本力, 鳥澤健太郎: 大規模情報分析システム WISDOM X, DISAANA, D-SUMM, 言語処理学会第 23 回年次大会発表論文集, pp. 1077-1080 (2017)
- [奎 13] 奎真奈見, 山本和英: 「やさしい日本語」変換システムの試作, 言語処理学会第 19 回年次大会発表論文集, pp. 678-681 (2013)
- [Moldovan 07a] Moldovan, D., Clark, C., and Bowden, M.: Lymba' s PowerAnswer 4 in TREC 2007, in the *Proceedings of the 16th Text REtrieval Conference*, pp. 1-9 (2007)
- [Moldovan 03] Moldovan, D., Clark, C., Harabagiu, S., and Maiorano, S.: COGEX: A logic prover for question answering, in *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 87-93 (2003)
- [Moldovan 07b] Moldovan, D., Clark, C., Harabagiu, S., and Hodges, D.: COGEX: A semantically and contextually enriched logic prover for question answering, in *Journal of Applied Logic*, Vol. 5, pp. 49-69 (2007)
- [Montano 13] Montano, R., Alias, F., and Ferrer, J.: Prosodic analysis of storytelling discourse modes and narrative situations oriented to text-to-speech synthesis, in *Proceedings of the 8th ISCA Speech Synthesis Workshop*, pp. 171-176 (2013)

- [Mori 04] Mori, T., Nozawa, M., and Asada, Y.: Multi-answer-focused multi-document summarization using a question-answering engine, in *Working Notes of the Fourth NTCIR Workshop Meeting*, pp. 422-429 (2004)
- [森本 04] 森本格行, 福本淳一: Why 型質問に対する回答抽出, 言語処理学会第 10 回年次大会発表論文集, pp. 293-296 (2004)
- [Morita 15] Morita, H., Kawahara, D., and Kurohashi, S.: Morphological Analysis for Unsegmented Languages using Recurrent Neural Network Language Model, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2292-2297 (2015)
- [森田 16] 森田一, 黒橋 禎夫: RNN 言語モデルを用いた日本語形態素解析の実用化, 情報処理学会第 78 回全国大会, pp. 13-14 (2016)
- [諸岡 06] 諸岡心, 福本淳一: Why 型質問応答のための回答選択手法, 電子情報通信学会技術研究報告, 言語理解とコミュニケーション, Vol. 105, No. 594, pp. 7-12 (2006)
- [諸岡 07] 諸岡心, 福本淳一: 非 Factoid 型質問に対応した質問応答システム, 言語処理学会第 13 回年次大会発表論文集, pp. 958-961 (2007)
- [Nakagawa 08] Nakagawa, T., Kawada, T., Inui, K., and Kurohashi, S.: Extracting subjective and objective evaluative expressions from the web, in *Proceedings of the 2nd International Symposium on Universal Communication*, pp. 251-258 (2008)
- [中川 09] 中川哲治, 乾健太郎, 黒橋禎夫: 隠れ変数を持つ条件付き確率場による依存構造木の評価極性分類, 情報処理学会研究報告, Vol. 2009-NL-192, No. 10, pp. 1-7 (2009)
- [Nakagawa 10] Nakagawa, T., Inui, K. and Kurohashi, S.: Dependency tree-based sentiment classification using CRFs with hidden variables, in *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 786-794 (2010)
- [Nallapati 17] Nallapati, R., Zhai, F., and Zhou, B.: SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents, in *Proceedings of the 31th AAAI Conference on Artificial Intelligence*, pp. 3075-3081 (2017)

- [Nallapati 16] Nallapati, R., Zhou, B., Santos, C.D., Gulcehre, C., and Xiang, B.: Abstractive text summarization using sequence-to-sequence RNNs and beyond, in *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pp. 280-290 (2016)
- [Napoles 10] Napoles, C. and Dredze, M.: Learning Simple Wikipedia: A cogitation in ascertaining abecedarian language, in *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids*, pp. 42-50 (2010)
- [Narisawa 13] Narisawa, K., Watanabe, Y., Mizuno, J., Okazaki, N., and Inui, K.: Is a 204 cm man tall or small ? Acquisition of numerical common sense from the web, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 382-391 (2013)
- [Ng 15] Ng, J.P. and Abrecht, V.: Better summarization evaluation with word embeddings for ROUGE, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1925-1930 (2015)
- [新名 17] 新名和也, 嶋田和孝: シズルワードから想起される料理の雑学を話す対話システム, 電子情報通信学会信学技報, Vol. 117, No. 367, NLC2017-42, pp. 77-82 (2017)
- [新名 18] 新名和也, 嶋田和孝: トリビア文抽出のためのトリビア度合いの推定, 電子情報通信学会信学技報, Vol. 118, No. 122, NLC2018-7, pp. 69-74 (2018)
- [西川 13] 西川仁, 平尾努, 牧野俊朗, 松尾義博, 松本裕治: 冗長性制約付きナップサック問題に基づく複数文書要約モデル, 自然言語処理, Vol. 20, No. 4, pp. 585-612 (2013)
- [西野 99] 西野文人, 橋本三奈子, 落谷亮: テキストからの用語とその定義文の抽出, 言語処理学会第5回年次大会発表論文集, pp. 124-127 (1999)
- [西野 12] 西野浩輝: 仕事ができる人の5日で身につく「伝える技術」, 東洋経済新報社 (2012)
- [西尾 14] 西尾泰和: word2vecによる自然言語処理, 株式会社オライリー・ジャパン, 初版第1刷 (2014)
- [西岡 09] 西岡晋太郎, 久保田裕章, 坂東晃文, 原田実: 意味グラフ照合による質問応答システム Metis の回答精度向上 - 質問文解析フェーズと検索フェーズの改良を中心として -, 研究報告自然言語処理 (NL) , Vol. 2009-NL-191, No. 16, pp. 1-8 (2009)

- [Nisimura 06] Nisimura, R., Omae, S., Kawahawa, H., and Irino, T.: Analyzing dialogue data for real-world emotional speech classification, in *Proceedings of the 7th Annual Conference of the International Speech Communication Association*, pp. 1822-1825 (2006)
- [Nomoto 08] Nomoto, T.: A generic sentence trimmer with CRFs, in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, Human Language Technologies*, pp. 299-307 (2008)
- [のざき 18] のざききこ: 声と言葉のプロが教える 伝わる話し方, 秀和システム (2018)
- [Oh 12] Oh, J.H., Torisawa, K., Hashimoto, C., Kawada, T., Saeger, S.D., Kazama, J., and Wang, Y.: Why question answering using sentiment analysis and word classes, in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 368-378 (2012)
- [Oh 13] Oh, J.H., Torisawa, K., Hashimoto, C., Sano, M., Saeger, S.D., and Ohtake, K.: Why-question answering using intra- and inter-sentential causal relations, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 1733-1743 (2013)
- [大泉 03] 大泉敏貴, 鍛冶伸裕, 河原大輔, 岡本雅史, 黒橋禎夫, 西田豊明: 書きことばから話しことばへの変換, 言語処理学会第9回年次大会発表論文集, pp. 93-96 (2003)
- [Okazaki 04] Okazaki, N., Matsuo, Y. and ishizuka, M.: TISS: An integrated summarization system for TSC-3, in *Working Notes of the 4th NTCIR Workshop Meeting*, pp. 436-443 (2004)
- [奥村 04] 奥村学: 特集 自然言語による情報アクセス技術, 3 テキスト自動要約, 情報処理, Vol. 45, No. 6, pp. 574-579 (2004)
- [奥村 05] 奥村学, 難波英嗣: テキスト自動要約, 知の科学, 人工知能学会編集, オーム社, 第1版 2刷 (2005)
- [奥村 13] 奥村学, 磯崎秀樹, 東中竜一郎, 永田昌明, 加藤恒昭: 質問応答システム, 自然言語処理シリーズ2, コロナ社, 初版第2刷 (2013)
- [大西 14] 大西可奈子, 吉村健: コンピュータとの自然な会話を実現する雑談対話技術, NTT DOCOMO テクニカル・ジャーナル, Vol. 21, No. 4, pp. 17-21 (2014)

- [小野 18] 小野喜己: ニューラルネットワークを用いたアクセント句境界及びアクセント核推定, 早稲田大学 2017 年度卒業論文, 第 2 章ニューラルネットワークを用いたアクセント推定, pp. 11-16 (2018)
- [Osborne 02] Osborne, M.: Using maximum entropy for sentence extraction, in *Proceedings of the Workshop on Automatic Summarization (including DUC 2002)*, pp. 1-8 (2002)
- [大田 09] 太田知宏, 鳥海不二夫, 石井健一郎: 発話生成を目的とした Wikipedia からの文抽出, 第 23 回人工知能学会全国大会論文集, 2G1-NFC5-11, pp. 1-4 (2009)
- [Paice 90] Paice, C.D.: Constructing literature abstracts by computer: Techniques and prospects, *Information Processing and Management*, Vol. 26, No. 1, pp. 171-186 (1990)
- [Palincsar 84] Palincsar, A.S. and Brown, A.L.: Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities, *Cognition and Instruction* 1, No. 2, pp. 117-175 (1984)
- [Pan 12] Pan, Y.C., Lee, H.Y., and Lee, L.S.: Interactive spoken document retrieval with suggested key terms ranked by a Markov decision process, *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 20, No. 2, pp. 632-645 (2012)
- [Pantel 06] Pantel, P. and Pennacchiotti, M.: Espresso : Leveraging generic patterns for automatically harvesting semantic relations, in *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pp. 113-120 (2006)
- [Papineni 02] Papineni, K., Roukos, S., Ward, T., and Zhu, W.J.: BLEU: A method for automatic evaluation of machine translation, in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311-318 (2002)
- [Paulus 18] Paulus, R., Xiong, C., and Socher, R.: A deep reinforced model for abstractive summarization, in *Proceedings of the 6th International Conference on Learning Representations*, pp. 1-12 (2018)
- [Pennington 14] Pennington, J., Socher, R., and Manning, C.D.: GloVe : Global vectors for word representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1532-1543 (2014)

- [Pichotta 16] Pichotta, K. and Mooney, R.J.: Using sentence-level LSTM language models for script inference, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 279-289 (2016)
- [Prakash 15] Prakash, A., Chinnakotla, M.K., Patel, D., and Garg, P.: Did you know? - Mining interesting trivia for entities from wikipedia, in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 3164-3170 (2015)
- [ProMarketing Wizard 13] ProMarketing Wizard: Trivia game on facebook, <http://bit.ly/1HUh1Rx> (2013)
- [Radev 97] Radev, D.R. and McKeown, K.R.: Building a generation knowledge source using Internet-accessible newswire, in *Proceedings of the fifth conference on Applied natural language processing*, pp. 221-228 (1997)
- [Rahman 16] Rahman, H., Hahn, T., and Segall, R.: Advanced feature-driven disease named entity recognition using conditional random fields, in *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 469-469 (2016)
- [Rajpurkar 16] Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P.: SQuAD: 100,000+ questions for machine comprehension of text, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2383-2392 (2016)
- [Rajpurkar 18] Rajpurkar, P., Jia, R., and Liang, P.: Know what you don't know: Unanswerable questions for SQuAD, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 784-789 (2018)
- [Ramli 16] Ramli, I., Seman, N., Ardi, N., and Jamil, N.: Rule-based storytelling text-to-speech (TTS) synthesis, in *Proceedings of the 3rd International Conference on Mechanics and Mechatronics Research*, pp 1-6 (2016)
- [Ramli 17] Ramli, I., Jamil, N., Seman, N., and Ardi, N.: The first Malay language storytelling text-to-speech (TTS) corpus for humanoid robot storytellers, *Journal of Fundamental and Applied Sciences*, Vol. 9, No. 4S, pp. 340-358 (2017)

- [Ranzato 16] Ranzato, M.A., Chopra, S., Auli, M., and Zaremba, W.: Sequence level training with recurrent neural networks, in *Proceedings of the International Conference on Learning Representations*, pp. 1-16 (2016)
- [Reithinger 00] Reithinger, N., Kipp, M., Engel, R., and Alexandersson, J.: Summarizing multilingual spoken negotiation dialogues, in *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pp. 1-8 (2000)
- [Rello 13] Rello, L., Baeza-Yates, R., Dempere-Marco, L., and Saggion, H.: Frequent words improve readability and short words improve understandability for people with dyslexia, *INTERACT 2013, Part IV, LNCS 8120*, Vol. 8120, pp. 203-219 (2013)
- [Rennie 17] Rennie, S.J., Marcheret, E., Mroueh, Y., Ross, J., and Goel, V.: Self-critical sequence training for image captioning, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7008-7024 (2017)
- [Rocktaschel 16] Rocktaschel, T., Grefenstette, E., Hermann, K.M., Kocisky, T., and Blunsom, P.: Reasoning about entailment with neural attention, in *Proceedings of the 4th International Conference on Learning Representations*, pp. 1-9 (2016)
- [Ronanki 16] Ronanki, S., Wu, Z., Watts, O., and King, S.: A demonstration of the Merlin open source neural network speech synthesis system, in *Proceedings of the 9th ISCA Speech Synthesis Workshop*, pp. 124-124 (2016)
- [Rush 15] Rush, A.M., Chopra, S., and Weston, J.: A neural attention model for sentence summarization, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 379-389 (2015)
- [Saeger 09a] Saeger, S.D., Torisawa, K., Kazama, J., Kuroda, K., and Murata, M.: Large scale relation acquisition using class dependent pattern, in *Proceedings of the IEEE International Conference on Data Mining*, pp. 764-769 (2009)
- [Saeger 09b] Saeger, S.D., Torisawa, K., Kazama, J., Kuroda, K., and Murata, M.: Large scale relation acquisition using class dependent patterns, in *Proceedings of the 2009 Ninth IEEE International Conference on Data Mining*, pp. 764-769 (2009)

- [Saeger 10] Stijn De Saeger, 鳥澤健太郎, 風間淳一, 黒田航, 村田真樹: 単語の意味クラスを用いたパターン学習による大規模な意味的關係獲得, 言語処理学会第 16 回年次大会発表論文集, pp. 932-935 (2010)
- [Saeger 12] Stijn De Saeger, 後藤淳, Istvan Varga: 音声質問応答システム「一休」, 情報通信研究機構季報, Vol.58, Nos.3/4, pp. 77-88 (2012)
- [貞光 12] 貞光九月, 齋藤邦子, 今村賢治, 松尾義博, 菊井玄一郎: トピック情報を用いたブーストラップ法に基づく語彙獲得, 自然言語処理, Vol. 19, No. 2, pp. 89-106 (2012)
- [貞光 13] 貞光九月, 東中竜一郎, 平野徹, 泉朋子: 知的な応答を実現するためのテキストからの知識獲得技術, NTT 技術ジャーナル, Vol. 20, No. 4, pp. 6-11 (2013)
- [桜井 02] 桜井裕, 佐藤理史: ワールドワイドウェブを利用した用語説明の自動生成, 情報処理学会論文誌, Vol. 43, No. 5, pp. 1470-1480 (2002)
- [Salton 96] Salton, G., Singhal, A., Buckley, C., and Mitra, M.: Automatic text decomposition using text segments and text themes, in *Proceedings of the 7th ACM Conference on Hypertext*, pp. 53-65 (1996)
- [Sankaran 16] Sankaran, B., Mi, H., Al-Onaizan, Y., Ittycheriah, A.: Temporal attention model for neural machine translation, in *arXiv:1608.02927*, pp. 1-8 (2016)
- [佐野 13] 佐野大樹, イシュトバーンヴァルガ, 鳥澤健太郎, 橋本力, 川田拓也, 呉鍾勲, 大竹清敬: 事象の活性化と不活性化を把握する言語資源の構築とその応用 - 災害時における問題報告と支援情報のマッチングを例に -, 第 4 回コーパス日本語学ワークショップ予稿集, pp. 203-212 (2013)
- [Sarkar 14] Sarkar, P., Haque, A., Dutta, A., Reddy, G., Harikrishna, M., Dhara, P., Verma, R., Narendra, P., Sunil, B., Yadav, J., and Rao, K.S.: Designing prosody rule-set for converting neutral TTS speech to storytelling style speech for indian languages: Bengali, Hindi and Telugu, in *Proceedings of the IEEE 7th International Conference on Contemporary Computing*, pp. 473-477 (2014)
- [Sarkar 15a] Sarkar, P. and Rao, K.S.: Modeling pauses for synthesis of storytelling style speech using unsupervised word features, *Procedia Computer Science*, Vol. 58, pp. 42-49 (2015)

- [Sarkar 15b] Sarkar, P. and Rao, K.S.: Data-driven pause prediction for speech synthesis in storytelling style speech, in *Proceedings of the 21st National Conference on Communications*, pp. 1-5 (2015)
- [Sarkar 15c] Sarkar, P. and Rao, K.S.: Data-driven pause prediction for synthesis of storytelling style speech based on discourse modes, *2015 IEEE International Conference on Electronics, Computing and Communication Technologies*, pp. 1-5 (2015)
- [佐々木 99] 佐々木裕: トランスデューサによる日本語固有表現抽出, 言語処理学会第 5 回年次大会発表論文集, pp. 108-111 (1999)
- [佐々木 01] 佐々木裕, 磯崎秀樹, 平博順, 平尾努, 賀沢秀人, 鈴木潤, 国領弘治, 前田英作: SAIQA : 大量文書に基づく質問応答システム, 情報処理学会研究報告, Vol. 86, pp. 77-82 (2001)
- [佐々木 04] 佐々木裕, 磯崎秀樹, 鈴木潤, 国領弘治, 平尾努, 賀沢秀人, 前田英作: SVM を用いた学習型質問応答システム SAIQA-II, 情報処理学会論文誌, Vol. 45, No. 2, pp. 635-646 (2004)
- [笹野 08] 笹野遼平, 黒橋禎夫: 大域的情報を用いた日本語固有表現認識, 情報処理学会論文誌, Vol. 49, No. 11, pp. 3765-3776 (2008)
- [Sasano 11a] Sasano, R. and Kurohashi, S.: A discriminative approach to Japanese zero anaphora resolution with large-scale lexicalized case frames, in *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pp. 758-766 (2011)
- [笹野 11b] 笹野遼平, 黒橋禎夫: 大規模格フレームを用いた識別モデルに基づく日本語ゼロ照応解析, 情報処理学会論文誌, Vol. 52, No. 12, pp. 3328-3337 (2011)
- [笹野 13] 笹野遼平, 河原大輔, 黒橋禎夫, 奥村学: 構文・述語項構造解析システム KNP の解析の流れと特徴, 言語処理学会第 19 回年次大会発表論文集, pp. 110-113 (2013)
- [佐藤 08] 佐藤理史: 日本語テキストの難易度判定ツール『帯』, Japio 2008 YEAR BOOK, 産業日本語特集, pp. 52-57 (2008)
- [Satt 17] Satt, A., Rozenberg, S., and Hoory, R.: Efficient emotion recognition from speech using deep learning on spectrograms, in *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, pp. 1089-1093 (2017)

- [Scarton 10] Scarton, C., de Oliveira, M., Candido Jr., A., Gasperin, C., and Alusio, S.M.: SIMPLIFICA: A tool for authoring simplified texts in Brazilian Portuguese guided by readability assessments, in *Proceedings of the NAACL HLT 2010 Demonstration Session*, pp. 41-44 (2010)
- [See 17] See, A., Liu, P.J., Manning, C.D.: Get to the point: Summarization with pointer-generator networks, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1073-1083 (2017)
- [Sekine 02] Sekine, S., Sudo, K., and Nobata, C.: Extended named entity hierarchy, in *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, pp. 1-7 (2002)
- [関根 07] 関根聡, 竹内康介: 拡張固有表現オントロジー, 言語処理学会第 13 回年次大会発表論文集, pp. 23-26 (2007)
- [Sekine 08] Sekine, S.: Extended named entity ontology, in *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pp. 52-57 (2008)
- [関根 18a] 関根聡, 安藤まや, 小林暁雄, 松田耕史, 鈴木正敏, Duc Nguyen, 乾健太郎: 「拡張固有表表現 + Wikipedia」データ, 言語処理学会第 24 回年次大会発表論文集, pp. 504-507 (2018)
- [関根 18b] 関根聡, 小林暁雄, 安藤まや, 馬場雪乃, 乾健太郎: Wikipedia 構造化データ「森羅」構築に向けて, 言語処理学会第 24 回年次大会発表論文集, pp. 765-768 (2018)
- [Selfridge 85] Selfridge, M., Daniell, J., and Simmons, D.: Learning causal models by understanding real-world natural language explanations, in *Proceedings of the 1st Conference on Artificial Intelligence for Applications*, pp. 378-383 (1985)
- [Seneff 00] Seneff, S. and Polifroni, J.: Dialogue management in the mercury flight reservation system, in *Proceedings of the 2000 ANLP/NAACL Workshop on Conversational Systems*, Vol. 3, pp. 11-16 (2000)
- [Serban 16] Serban, I.V., Sordoni, A., Bengio, Y., Courville, A., and Pineau, J.: Building end-to-end dialogue systems using generative hierarchical neural network models, in *Proceedings of the 30th AAAI Conference on Artificial Intelligence, Association for the Advancement in Artificial Intelligence*, pp. 3776-3783 (2016)

- [Shang 15] Shang, L., Lu, Z., and Li, H.: Neural responding machine for short-text conversation, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 1577-1586 (2015)
- [Shardlow 14] Shardlow, M.: A survey of automated text simplification, *International Journal of Advanced Computer Science and Applications, Special Issue on Natural Language Processing 2014*, pp. 58-70 (2014)
- [Sharp 16] Sharp, R., Surdeanu, M., Jansen, P., Clark, P., and Hammond, M.: Creating causal embeddings for question answering with minimal supervision, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 138-148 (2016)
- [渋沢 07] 渋沢潮, 林貴宏, 尾内理紀夫: Why 型質問の回答文を WEB から自動抽出するシステムの開発と評価, 情報処理学会論文誌, Vol. 48, No. 3, pp. 1512-1523 (2007)
- [しまね国際センター] やさしい日本語の手引, 島根県・(公財) しまね国際センター, <http://www.sic-info.org/sic/post-1318/>
- [Shinzato 08] Shinzato, K., Shibata, T., Kawahara, D., Hashimoto, C., and Kurohashi, S.: Tsubaki: An open search engine infrastructure for developing new information access: in *Proceedings of the 3rd International Joint Conference on Natural Language Processing*, pp. 189-196 (2008)
- [Siddharthan 04] Siddharthan, A., Nenkova, A., and Mckeown, K.: Syntactic simplification for improving content selection in multi-document summarization, in *Proceedings of the 20th International Conference on Computational Linguistics*, pp. 896-902 (2004)
- [Silveira 12] Silveira, S. and Branco, A.: Combining a double clustering approach with sentence simplification to produce highly informative multi-document summaries, *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference*, pp. 482-489 (2012)
- [Skorokhod'ko 72] Skorokhod'ko, E.F.: Adaptive method of automatic abstracting and indexing, *Information Processing* 71, pp. 1179-1182 (1972)

- [Sohrab 18] Sohrab, M.G. and Miwa, M.: Deep exhaustive model for nested named entity recognition, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2843-2849 (2018)
- [Sordoni 15a] Sordoni, A., Galley, M., Auli, M., Brockett, C., Mitchell, M., Nie, J.Y., Gao, J., and Dolan, B.: A neural network approach to context-sensitive generation of conversational responses, in *Proceedings of the Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, pp. 196-205 (2015)
- [Sordoni 15b] Sordoni, A., Bengio, Y., Vahabi, H., Lioma, C., Simonsen, J.G., and Nie, J.Y.: A hierarchical recurrent encoder-decoder for generative context-aware query suggestion, in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 553-562 (2015)
- [Soricut 06] Soricut, R. and Brill, E.: Automatic question answering using the web: Beyond the factoid, *Journal of Information Retrieval*, Vol. 9, No. 2, pp. 191-206 (2006)
- [Specia 10] Specia, L.: Translating from complex to simplified sentences, in *Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language*, pp. 30-39 (2010)
- [Srivastava 15] Srivastava, R.K., Greff, K., and Schmidhuber, J.: Highway networks, in *Proceedings of the Deep Learning Workshop in the 32nd International Conference on Machine Learning*, pp. 1-6 (2015)
- [Stajner 15] Stajner, S., Bechara, H., and Saggion, H.: A deeper exploration of the standard PB-SMT approach to text simplification and its evaluation, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 823-828 (2015)
- [Sugiura 14] Sugiura, K., Shiga, Y., Kawai, H., Misu, T., and Hori, C.: Non-monologue HMM-based speech synthesis for service robots: A cloud robotics approach, *IEEE International Conference on Robotics and Automation*, pp. 2237-2242 (2014)
- [Sugiyama 00] 杉山聡, 堂坂浩二, 川端豪: 音声対話によるテキスト内容の伝達方法, 情処学論, Vol. 41, No. 6, pp. 1883-1894 (2000)

- [砂川 12a] 砂川有里子: 学習辞書編集支援データベース作成について - 『学習辞書科研』プロジェクトの紹介, 日本語教育連絡会議論文集, Vol. 24, pp. 164-169 (2012)
- [Sunakawa 12b] Sunakawa, Y., Lee, J., and Takahara, M.: The construction of a database to support the compilation of Japanese learners' dictionaries, *Acta Linguistica Asiatica*, Vol. 2, No. 2, pp. 97-115 (2012)
- [Sutskever 14] Sutskever, I., Vinyals, O., and Le, Q.V.: Sequence to sequence learning with neural networks, in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Vol. 2, pp. 3104-3112 (2014)
- [諏訪 04] 諏訪善彦, 鳥澤健太郎: EM 法による単語意味分類のための効果的セッティング, 言語処理学会第 10 回年次大会発表論文集, pp. 1-4 (2004)
- [Suzuki 02] Suzuki, J., Sasaki, Y., and Maeda, E.: SVM answer selection for open-domain question answering, in *Proceedings of the 19th international conference on Computational linguistics*, pp. 1-7 (2002)
- [鈴木 03] 鈴木潤, 佐々木裕, 前田英作, 単語属性 N-gram と統計的機械学習による質問タイプ同定, 情報処理学会論文誌, Vol. 44, No. 11 (2003)
- [鈴木 18] 鈴木正敏, 松田耕史, 岡崎直観, 乾健太郎: 読解による解答可能性を付与した質問応答データセットの構築, 言語処理学会第 24 回年次大会発表論文集, pp. 702-705 (2018)
- [Suzuki 17] Suzuki, J. and Nagata, M.: Cutting-off redundant repeating generations for neural abstractive summarization, in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 291-297 (2017)
- [Szpektor 08] Szpektor, I. and Dagan, I.: Learning entailment rules for unary templates, in *Proceedings of the 22nd International Conference on Computational Linguistics*, pp. 849-856 (2008)
- [高木 10] 高木想一郎, 飯田龍, 徳永健伸: 照応関係を利用した日本語質問文の生成, 言語処理学会第 16 回年次大会発表論文集, pp. 816-819 (2010)
- [Takamura 05a] Takamura, H., Inui, T., and Okumura, M.: Extracting semantic orientations of words using spin model, in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 133-140 (2005)

- [Takamura 05a] 高村大也, 乾孝司, 奥村学: スピンモデルの感情情報処理への応用, 言語処理学会第 11 回年次大会, pp. 648-651 (2005)
- [Takamura 06] 高村大也, 乾孝司, 奥村学: スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌, Vol. 47, No. 2, pp. 627-637 (2006)
- [Takamura 09] Takamura, H. and Okumura, M.: Text summarization model based on maximum coverage problem and its variant, in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 781-789 (2009)
- [高津 15] 高津弘明, 小林哲則: 対話エージェントのための性格モデル, 言語処理学会第 21 回年次大会発表論文集, pp. 191-194 (2015)
- [高津 16] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための口語化における述語の書き換え, 言語処理学会第 22 回年次大会発表論文集, pp. 1033-1036 (2016)
- [高山 10] 高山真行, 今村泰香, 久保田裕章, 原田実: NonFactoid 型質問文と回答文との意味的關係に基づく質問応答システム Metis, 研究報告音声言語情報処理 (SLP), Vol. 2010-SLP-81, No. 17, pp. 1-7 (2010)
- [武田 91a] 武田昌一, 市川熹: 日本語文音声におけるプロミネンスの韻律的特徴の解析, 日本音響学会誌, Vol. 47, No. 6, pp. 386-396 (1991)
- [武田 91b] 武田昌一, 市川熹: 日本語文音声のプロミネンス生成規則の作成と評価, 日本音響学会誌, Vol. 47, No. 6, pp. 397-404 (1991)
- [竹内 10] 竹内孔一, 土山傑, 守屋将人, 森安祐樹: 類似した動作や状況を検索するための意味役割及び動詞語義付与システムの構築, 言語理解とコミュニケーション研究会, Vol. IEICE-109, No. 390, pp. 1-6 (2010)
- [Tan 16] Tan, M., Santos, C., Xiang, B., and Zhou, B.: Improved representation learning for question answer matching, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 464-473 (2016)
- [田中 98] 田中真詞, 川端豪: 文型と音調によるユーザの発話意図の推定, 情報処理学会研究報告音声言語情報処理 (SLP), Vol. 1998, No. 68(1998-SLP-022), pp. 55-60 (1998)

- [田中 10] 田中英輝, 美野秀弥: やさしい日本語によるニュースの書き換え実験, 情報処理学会研究報告, Vol. 2010-NL-199, No. 11, pp. 1-8 (2010)
- [田中 13] 田中英輝, 美野秀弥, 越智慎司, 柴田元也, やさしい日本語ニュースの公開実験, NHK 技研 R&D, No. 139, pp. 20-29 (2013)
- [Tanaka 13] Tanaka, M., Saeger, S.D., Ohtake, K., Hashimoto, C., Hijiya, M., Fujii, H., and Torisawa, K.: WISDOM2013 : A large-scale web information analysis system, in *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pp. 45-48, 2013.
- [田仲 14] 田仲正弘, 大竹清敬, 鳥澤健太郎, 田浦健次郎: RaSC : 高速なストリーム通信をサポートする言語処理プログラムの高速化・高並列化ミドルウェア, 言語処理学会第 20 回年次大会発表論文集, pp. 705-708 (2014)
- [Tang 18] Tang, D., Zeng, J., and Li, M.: An end-to-end deep learning framework for speech emotion recognition of atypical individuals, in *Proceedings of the 19th Annual Conference of the International Speech Communication Association*, pp. 162-166 (2018)
- [Tatu 07] Tatu, M. and Moldovan, D.: COGEX at RTE3, in *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pp. 22-27 (2007)
- [Theune 06] Theune, M., Meijs, K., Heylen, D., and Ordelman, R.: Generating expressive speech for storytelling applications, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 4, pp. 1137-1144 (2006)
- [Thomas 12] Thomas, S.R. and Anderson, S.: WordNet-based lexical simplification of a document, in *Proceedings of KONVENS 2012*, pp. 80-88 (2012)
- [Tolmachev 18a] Tolmachev, A. and Kurohashi, S.: Juman++ v2: A practical and modern morphological analyzer, 言語処理学会第 24 回年次大会発表論文集, pp. 917-920 (2018)
- [Tolmachev 18b] Tolmachev, A., Kawahara, D., and Kurohashi, S.: Juman++: A morphological analysis toolkit for scriptio continua, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (System Demonstrations)*, pp. 54-59 (2018)

- [Torisawa 01] Torisawa, K.: An unsupervised method for canonicalization of Japanese post-positions, in *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium*, pp. 211-218 (2001)
- [Tran 18] Tran, N.K. and Niedereese, C.: Multihop attention networks for question answer matching, in *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 325-334 (2018)
- [Traum 15a] Traum, D., Georgila, K., Artstein, R., and Leuski, A.: Evaluating spoken dialogue processing for time-offset interaction, in *Proceedings of the 16th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 199-208 (2015)
- [Traum 15b] Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., Haase, K., Jungblut, K., Leuski, A., Smith, S., and Swartout, W.: New Dimensions in Testimony: Digitally Preserving a Holocaust Survivor's Interactive Storytelling, *Interactive Storytelling: 8th International Conference on Interactive Digital Storytelling*, pp. 269-281 (2015)
- [Traum 16] Traum, D.: 双方向型歴史学習の支援のための対話システム技術の活用, 人工知能学会誌, Vol. 31, No. 6, pp. 806-812 (2016)
- [Trischler 17] Trischler, A., Wang, T., Yuan, X., Harris, J., Sordoni, A., Bachman, P., and Suleman, K.: NewsQA : A machine comprehension dataset, in *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pp. 191-200 (2017)
- [土田 04] 土田正明, 松井藤五郎, 大和田勇人: World Wide Web を用いた辞典システムの構築, 人工知能学会第 18 回全国大会論文集, pp. 1-3 (2004)
- [Tsuchida 11a] Tsuchida, M., Torisawa, K., Saeger, S.D., Oh, J.H., Kazama, J., Hashimoto, C., and Ohwada, H.: Toward finding semantic relations not written in a single sentence: An inference method using auto-discovered rules, in *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pp. 902-910 (2011)
- [土田 11b] 土田正明, 鳥澤健太郎, De Saeger Stijn, 呉鍾勲, 風間淳一, 大和田勇人: 推論ルール学習による単語間の意味的關係推論法, 語処理学会第 17 回年次大会発表論文集, pp. 81-84 (2011)

- [Tsurel 17] Tsurel, D., Pelleg, D., Guy, I., and Shahaf, D.: Fun facts: Automatic trivia fact extraction from wikipedia, in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 345-354 (2017)
- [内田 13] 内田 渉, 森田千晶, 吉村健: 自然文質問への直接回答を実現する知識 Q&A, NTT DO-COMO テクニカル・ジャーナル, Vol. 20, No. 4, pp. 21-24 (2013)
- [梅澤 08] 梅澤 俊之, 西尾 華織, 松田 源立, 原田 実: 意味解析システム SAGE の精度向上とモダリティの付与と辞書更新支援系の開発, 言語処理学会第 14 回年次大会発表論文集, pp. 548-551 (2008)
- [Vaswani 17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., and Kaiser, L.: Attention is all you need, in *Proceedings of the 31st Conference on Neural Information Processing System*, pp. 6000-6010 (2017)
- [Verma 15] Verma, R., Sarkar, P., and Rao, K.S.: Conversion of neutral speech to storytelling style speech, in *Proceedings of the 8th International Conference on Advances in Pattern Recognition*, pp. 1-6 (2015)
- [Vilain 96] Vilain, M. and Day, D.: Finite-state phrase parsing by rule sequences, in *Proceedings of the 16th International Conference on Computational Linguistics* pp. 274-279 (1996)
- [Vinyals 15a] Vinyals, O., Fortunato, M., and Jaitly, N.: Pointer networks, in *Advances in Neural Information Processing Systems* 28, pp.2692-2700, 2015.
- [Vinyals 15b] Vinyals, O. and Le, Q.V.: A neural conversational model, in *Proceedings of the 31st International Conference on Machine Learning*, pp. 1-8 (2015)
- [Voices Heard Media 13] Voices Heard Media: Using trivia and quiz products to engage your customer, <http://bit.ly/1Htiv5V> (2013)
- [Voorhees 01] Voorhees, E.M.: Question answering in TREC, in *Proceedings of the 10th International Conference on Information and Knowledge Management*, pp. 535-537 (2001)
- [Wang 16] Wang, S. and Jiang, J.: Learning natural language inference with LSTM, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1442-1451 (2016)

- [Wang 16b] Wang, X. and Wang, D.: An implementation of Pointer-Networks with Extensions, *CSLT TECHNICAL REPORT-20150031*, pp. 1-7 (2016)
- [Wang 18] Wang, W., Yan, M., and Wu, C.: Multi-granularity hierarchical attention fusion networks for reading comprehension and question answering, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 1705-1714 (2018)
- [Watanabe 96] Watanabe, H.: A method for abstracting newspaper articles by using surface clues, in *Proceedings of the 16th International Conference on Computational Linguistics*, pp. 974-979 (1996)
- [Weeds 03] Weeds, J. and Weir, D.: A general framework for distributional similarity, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 81-88 (2003)
- [Weissenborn 17] Weissenborn, D., Wiese, G., and Seiffe, L.: Making neural QA as simple as possible but not simpler, in *Proceedings of the 21st Conference on Computational Natural Language Learning*, pp. 271-280 (2017)
- [Wilcock 12] Wilcock, G.: WikiTalk: A Spoken Wikipedia-based Open-Domain Knowledge Access System, in *Proceedings of the COLING 2012 Workshop on Question Answering for Complex Domains*, pp. 57-69 (2012)
- [Wilcock 13] Wilcock, G. and Jokinen, K.: WikiTalk Human-Robot Interactions, in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, pp. 73-74 (2013)
- [Wilcock 15] Wilcock, G. and Jokinen, K.: Multilingual WikiTalk: Wikipedia-based talking robots that switch languages, in *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 162-164 (2015)
- [Wu 15] Wu, Z., Swietojanski, P., Veaux, C., Renals, S., and King, S.: A study of speaker adaptation for DNN-based speech synthesis, in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, pp. 879-883 (2015)
- [Wu 16] Wu, Z., Watts, O., and King, S.: Merlin: An open source neural network speech synthesis system, in *Proceedings of the 9th ISCA Speech Synthesis Workshop*, pp. 218-223 (2016)

- [Wubben 12] Wubben, S., van den Bosch, A., and Krahmer, E.: Sentence simplification by monolingual machine translation, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pp. 1015-1024 (2012)
- [Xu 03] Xu, J., Licuanan, A., and Weischedel, R.: TREC 2003 QA at BBN: Answering definitional questions, in *Proceedings of the 12th Text REtrieval Conference*, pp. 1-9 (2003)
- [Xu 05] Xu, J., Cao, Y., Li, H., and Zhao, M.: Ranking definitions with supervised learning methods, in *Proceedings of the 14th International Conference on World Wide Web*, pp. 811-819 (2005)
- [Xu 16] Xu, W., Napoles, C., Pavlick, E., Chen, Q., and Callison-Burch, C.: Optimizing statistical machine translation for text simplification, *Transactions of the Association for Computational Linguistics*, Vol. 4, pp. 401-415 (2016)
- [Yadav 18] Yadav, V. and Bethard, S: A survey on recent advances in named entity recognition from deep learning models, in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 2145-2158 (2018)
- [山村 15] 山村崇, 陽徳永, 嶋田和孝: 時間情報とテキストセグメンテーションに基づく複数人対話要約手法, 電子情報通信学会技術研究報告, Vol. 115, No. 70, pp. 43-48 (2015)
- [山本 17] 本和英, 丸拓海, 張晴, 稲岡夢, 川耀朗, 勝哲弘, 高橋寛治: やさしい本語対訳コーパスの構築, 言語処理学会第 23 回年次大会発表論文集, pp. 763-766 (2017)
- [Yan 12] Yan, Y., Hashimoto, C., and Torisawa, K.: Pattern mining approach to unsupervised definition extraction, 言語処理学会第 18 回年次大会発表論文集, pp. 459-462 (2012)
- [Yang 03] Yang, H., Cui, H., Maslennikov, M., Qiu, L., Kan, M.Y., and Chua, T.S.: QUALIFIER in TREC-12 QA main task, in *Proceedings of the 12th Text REtrieval Conference*, pp. 1-9 (2003)
- [Yatskar 10] Yatskar, M., Pang, B., Danescu-Niculescu-Mizil, C., and Lee, L.: For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia, in *Proceedings of the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 365-368 (2010)

- [Yenigalla 18] Yenigalla, P., Kumar, A., Tripathi, S., Singh, C., Kar, S., and Vepa, J.: Speech emotion recognition using spectrogram & phoneme embedding, in *Proceedings of the 19th Annual Conference of the International Speech Communication Association*, pp. 3688-3692 (2018)
- [Yih 07] Yih, W., Goodman, J., Vanderwende, L., and Suzuki, H.: Multi-document summarization by maximizing informative content-words, in *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pp. 1776-1782 (2007)
- [横手 13] 横手健一, ボレガラ・ダヌシカ, 石塚 満: テキスト含意認識に有効な意味類似度変換及びその獲得法, 人工知能学会誌, Vol. 28, No. 2, pp. 220-229 (2013)
- [Yokoyama 18] Yokoyama, K., Takatsu, H., Honda, H., Fujie, S., and Kobayashi, T.: Investigation of users' short responses in actual conversation system and automatic recognition of their intentions, in *Proceedings of the the 2018 IEEE Workshop on Spoken Language Technology* (2018)
- [Yoshinaga 12] Yoshinaga, N. and Kitsuregawa, M.: Efficient classification with conjunctive features, *Journal of Information Processing*, Vol. 20 No. 1 pp. 228-237 (2012)
- [Yoshinaga 14] Yoshinaga, N. and Kitsuregawa, M.: A self-adaptive classifier for efficient text-stream processing, in *Proceedings of the 24th International Conference on Computational Linguistics*, pp. 1091-1102 (2014)
- [吉野 11] 吉野幸一郎, 森信介, 河原達也: 述語項の類似度に基づく情報抽出・推薦を行う音声対話システム, 情報処理学会論文誌, Vol. 52, No. 12, pp. 3386-3397 (2011)
- [吉野 14] 吉野幸一郎, 河原達也: ユーザの焦点を用いた POMDP による音声情報案内システム, 人工知能学会言語・音声理解と対話処理研究会, 2014-SLUD72, pp. 1-6 (2014)
- [Yoshino 15a] Yoshino, K. and Kawahara, T.: Conversational system for information navigation based on POMDP with user focus tracking, *Computer Speech & Language*, Vol. 34, No. 1, pp. 275-291 (2015)
- [Yoshino 15b] Yoshino, K. and Kawahara, T.: News navigation system based on proactive dialogue strategy, in *Proceedings of the International Workshop series on Spoken Dialogue Systems Technology*, pp. 1-11 (2015)

- [Yu 18] Yu, A.W., Dohan, D., Luong, M.T., Zhao, R., Chen, K., Norouzi, M., and Le, Q.V.: QANet: Combining local convolution with global self-attention for reading comprehension, in *Proceedings of the 6th International Conference on Learning Representations*, pp. 1-16 (2018)
- [Zechner 96] Zechner, K.: Fast generation of abstracts from general domain text corpora by extracting relevant sentences, in *Proceedings of the 16th International Conference on Computational Linguistics*, pp. 986-989 (1996)
- [Zechner 01] Zechner, K., and Lavie, A.: Increasing the coherence of spoken dialogue summaries by cross-speaker information linking, in *Proceedings of the NAACL2001 Workshop on Automatic Summarization*, pp. 22-31 (2001)
- [Zen 13] Zen, H., Senior, A., and Schuster, M.: Statistical parametric speech synthesis using deep neural networks, in *Proceedings of the 2013 IEEE International Conference on Acoustics Speech and Signal Processing*, pp. 7962-7966 (2013)
- [Zhao 14] Zhao, R., Papangelis, A., and Cassell, J.: Towards a dyadic computational model of rapport management for human-virtual agent interaction, in *Proceedings of the International Conference on Intelligent Virtual Agents*, pp. 514-527 (2014)
- [Zhao 16] Zhao, Y., Saito, D., and Minematsu, N.: Speaker representations for speaker adaptation in multiple speakers' BLSTM-RNN-based speech synthesis, in *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, pp. 2268-2272 (2016)
- [Zhou 06] Zhou, Y., Yuan, X., Cao, J., Huang, X., Wu, L.: FDUQA on TREC 2006 QA track, in *Proceedings of the 15th Text REtrieval Conference*, pp. 1-10 (2006)
- [Zhu 10] Zhu, Z., Bernhard, D., and Gurevych, I.: A monolingual tree-based translation model for sentence simplification, in *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 1353-1361 (2010)

研究業績

論文

- [1] 高津弘明, 福岡維新, 藤江真也, 岩田和彦, 小林哲則: 会話によるニュース記事伝達のための音声合成, 人工知能学会論文誌, Vol. 34, No. 2, pp. 1-15 (2019)
- [2] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 意図性の異なる多様な情報行動を可能とする音声対話システム, 人工知能学会論文誌, Vol. 33, No. 1, pp. 1-24 (2018)

講演

- [3] 高津弘明, 横山勝矢, 本田裕, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための発話意図理解, 人工知能学会第32回全国大会論文集, 4Pin1-29, pp. 1-4 (2018)
- [4] 高津弘明, 横山勝矢, 本田裕, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための間の調整, 言語処理学会第24回年次大会発表論文集, pp. 212-215 (2018)
- [5] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための文圧縮, 言語処理学会第23回年次大会発表論文集, pp. 154-157 (2017)
- [6] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 意図性の異なる多様な情報行動を可能とする音声対話システム, 人工知能学会第30回全国大会論文集, 2O4-OS-23a-3in2, pp. 1-4 (2016)
- [7] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 快適な情報享受を可能とする音声対話システム, 言語処理学会第22回年次大会発表論文集, pp. 302-305 (2016)
- [8] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための口語化における述語の書き換え, 言語処理学会第22回年次大会発表論文集, pp. 1033-1036 (2016)
- [9] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための情報選択, 言語処理学会第22回年次大会発表論文集, pp. 199-202 (2016)

- [10] 高津弘明, 小林哲則: 多人数会話における議論支援のための状態把握, 言語処理学会第 22 回年次大会発表論文集, pp. 497-500 (2016)
- [11] 高津弘明, 小林哲則: 分散表現を用いた動詞・フレーズの含意関係認識, 言語処理学会第 21 回年次大会発表論文集, pp. 928-931 (2015)
- [12] 高津弘明, 小林哲則: 対話エージェントのための性格モデル, 言語処理学会第 21 回年次大会発表論文集, pp. 191-194 (2015)

共著

- [13] Katsuya Yokoyama, Hiroaki Takatsu, Hiroshi Honda, Shinya Fujie, and Tetsunori Kobayashi: Investigation of users' short responses in actual conversation system and automatic recognition of their intentions, in *Proceedings of the 2018 IEEE Workshop on Spoken Language Technology* (2018)
- [14] 横山勝矢, 高津弘明, 本田裕, 藤江真也, 小林哲則: 会話によるニュース記事伝達のための発話意図の分類と認識, 第 123 回音声言語情報処理研究会, pp. 1-8 (2018)
- [15] 横山勝矢, 高津弘明, 本田裕, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための発話意図分類とデータベースの構築, 人工知能学会第 32 回全国大会論文集, 3Pin1-23, pp. 1-4 (2018)
- [16] 福岡維新, 高津弘明, 藤江真也, 林良彦, 小林哲則: 会話による情報伝達における発話系列の韻律分析, 人工知能学会第 30 回全国大会論文集, 2P1-5in1, pp. 1-4 (2016)
- [17] Shinya Fujie, Ishin Fukuoka, Asumi Mugita, Hiroaki Takatsu, Yoshihiko Hayashi, and Tetsunori Kobayashi: A spoken dialog system for coordinating information consumption and exploration, in *Proceedings of the 1st ACM SIGIR Conference on Human Information Interaction and Retrieval*, pp. 253-256 (2016)
- [18] 林良彦, 藤江真也, 福岡維新, 高津弘明, 小林哲則: 情報アクセスにおける受動性と能動性: 音声対話によるニュース記事アクセス, 人工知能学会第 11 回インタラクティブ情報アクセスと可視化マイニング研究会, pp. 25-32 (2015) (研究会優秀賞受賞)
- [19] 藤江真也, 福岡維新, 麥田愛純, 高津弘明, 林良彦, 小林哲則: 聞き手の反応を考慮した発話計画を用いた音声対話システム, 日本音響学会 2015 年秋季研究発表会, pp. 1-4 (2015)

- [20] 藤江真也, 福岡維新, 麥田愛純, 高津弘明, 林良彦, 小林哲則: 効率的な情報伝達を指向した音声対話システムの提案, 人工知能学会第 74 回言語・音声理解と対話処理研究会, SIG-SLUD-B501-02, pp. 7-12 (2015)
- [21] 福岡維新, 麥田愛純, 高津弘明, 藤江真也, 林良彦, 小林哲則: 漸増的な情報補完機能を有する音声対話システム, 人工知能学会第 29 回全国大会論文集, 2L4-OS-07a, pp. 1-4 (2015)