

『日本語日常会話コーパス』における 談話行為タグの利用可能性

2021年3月4日
シンポジウム「日常会話コーパス」VI

居間 友里子 (国立国語研究所)

2021/3/4

1

はじめに

- ▶ 本発表で注目するもの：**談話行為情報**
 - ▶ 会話内でなされた各発話で行われている行為や談話の管理に関する機能などに関する情報
 - ▶ 『日本語日常会話コーパス (CEJC) 』にも、この情報に関するタグである「談話行為タグ」を付与、公開予定
 - ▶ 検索のための使用を想定
- ▶ 本発表の概要
 - ▶ 談話行為タグがどのようなものか紹介
 - ▶ どういった利用可能性を有しているのかについて紹介

2

はじめに

- ▶ CEJC (200時間)
 - ▶ [人手作成] 転記テキスト
 - ▶ [自動付与] 単語・係り受け・発話単位
- ▶ コア (20時間)
 - ▶ + [人手修正] 単語・係り受け・発話単位
 - ▶ + [人手付与] 韻律情報・**談話行為情報**

3

はじめに

▶ 転記の例

会話ID	開始	終了	話者	発話	談話行為情報
C002_003	5.252	6.142	IC01_日野	何時に出んの?。	🗨️
C002_003	9.78	12.167	IC02_智也	だから(0.571)用事が終わったら出るけど。	🗨️

- ▶ ここに**談話行為情報**を追加

4

本発表の流れ

1. CEJCに付与している談話行為タグの紹介
 - ▶ 談話行為タグにかんする先行研究
 - ▶ CEJCの採用している枠組み
2. 付与したタグから見えること
 - ▶ CEJCコアにおける談話行為の分布
 - ▶ 談話行為タグと会話の特徴
 - ▶ 当該会話の特徴は談話行為に反映されるのか
3. 今後の課題と展望
 - ▶ 談話行為タグが扱えるもの・扱えないもの

5

先行研究

- ▶ 談話行為情報
 - ▶ 会話内でなされた各発話の際になされている行為や談話の管理に関する機能などに関する情報
 - ▶ 対話システムや自動要約システムの構築などへの応用を視野
- ▶ 会話データに談話行為情報を付与する試み
 - ▶ 国内外で広く行われている (Allen and Core, 1997; Jurafsky et al., 1997など) (荒木他 1999; 人工知能学会 談話・対話研究におけるコーパス利用研究グループ2000 など)
 - ▶ 国際標準化規格 (Bunt et al., 2012; ISO 24617-2, 2012)
 - ▶ 相互行為の調整に関する機能も扱える
 - ▶ 重層的に複数の談話機能を付与できる
 - ▶ 「会話は発話を通して複数のことがなされる多面的な活動」 (Allwood, 1977)
 - ▶ 例) えーと Aちゃん 明日行く？

6

CEJCの性質と談話行為タグ

- ▶ CEJCの特徴
 - ▶ 対面で行われた会話を中心 (=参加者たちは視覚的資源も使用できる)
 - ▶ 日常に埋め込まれた多様な活動を扱っている
- ▶ 談話行為タグ付与の視点からみると、
 - ▶ 非タスク志向の会話場面や会話に集中していない状況も多い
 - = 話題や行為が曖昧でタグ判断が困難な発話も多い
 - ▶ 一つの枠組みを多様な場面に適用する
 - = 各会話の特性にアノテーション枠組みをびったり沿わせることができない
 - = 会話間の特性の比較が可能

7

CEJCでの談話行為タグ

- ▶ CEJCに談話行為情報を付与する目的
 - ▶ 特定行為やその連なりが生じている箇所をデータから見つけ出すために使用
 - ▶ やり取りの骨組みとなる特徴が安定的に示されることをめざす
- ▶ 採用した枠組み：国際標準規格ISO24617-2 を参考に一部変更
 - ▶ 単位：発話単位 (LUU: Japanese Discourse Research Initiative, 2014)
 - ▶ 付与する情報：二段階で付与 (居間・伝, 2018) →多機能性を反映可能
 - ▶ 第一レベル：基本的な談話機能
 - ▶ 第二レベル：談話の展開やより大局的な行為連鎖に関わる情報/相互行為の調整
 - ▶ 依存関係：発話同士の結びつき方
- ▶ 付与対象：CEJCコア (20時間)

8

第一レベルタグ

- ▶ すべての発話単位に付与する
 - ▶ 基本的な談話機能を扱うもの

タスク系	情報提供/情報要求/返答としての情報・確認提供/ 依頼系(依頼/指示/命令/勧誘/提案)/依頼系への対処/申し出/ 申し出への対処/注意獲得/注意獲得への対処/独り言
フィードバック系	肯定/了承/反復/語彙的反応
社会的付き合い管理系	挨拶/謝罪/謝罪への対処/感謝/感謝への対処
その他	非言語行動/合図/不明/対象外

9

第二レベルタグ

- ▶ 該当する場合のみ付与する
 - ▶ 談話の展開やより大局的な行為連鎖に関わる情報を扱うもの
 - ▶ 相互行為の調整に関わる情報を扱うもの

準備系	準備の準備/準備/準備系への対処/準備が投射する本体
回収系	回収/回収への対処/補完/補完への対処
修復系	修復開始/修復操作
談話構造化系	談話開始準備/談話開始/談話開始系への対処/談話終了準備/ 談話終了/談話終了系への対処
順番管理系	順番取得/順番開放/順番維持/順番放棄
メタ発話情報系	読み上げ/冗談

10

依存関係タグ

- ▶ 依存関係情報
 - ▶ 当該発話が他の発話とどのような関係で結ばれているかにかんする情報
- ▶ 依存関係タグ
 1. 予測的依存関係 挨拶-挨拶/質問-応答/依頼-受け入れ・拒否....
 2. 適及的依存関係 情報提供-フィードバック/質問-応答-フィードバック....

+

 - ▶ 依存先の発話を指定

11

依存関係タグ

発話ID	話者	発話	LV1	LV1 依存関係	LV1 依存先	LV2	LV2 依存関係	LV2 依存先
2	IC01_日野	何時に出んの?。	T_情報要求					
3	IC02_智也	うん?。	T_情報要求			修復開始	適及的	2
4	IC02_智也	きょう?。	T_情報要求			修復開始	適及的	2
5	IC01_日野	うん。	T_返答	予測的	4	修復	予測的	4
6	IC02_智也	だから(0.571)用事が 終わったら出るけど。	T_返答	予測的	2			

12

CEJCでの談話行為タグ

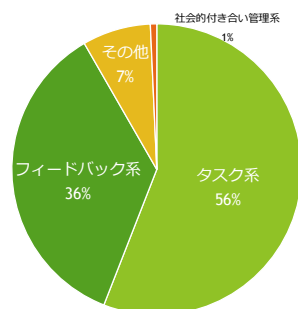
- ▶ これらのタグを用いて
 - ▶ 特定の行為を行っている発話箇所を探す
 - ▶ 申し出が行われている箇所/謝罪が行われている箇所
 - ...特定の言語形式上の特徴を持たなくても拾い出すことができる
 - ▶ 特定の行為の連なりが生じている箇所を探す
 - ▶ 談話行為 x 依存関係情報 x 参照先情報 を組み合わせる
 - ▶ 提案に提案を返すような発話
 - ▶ 呼び出しに応答が返される例、返されない例
 - ▶ 修復の連鎖
 - ▶ 他のアノテーションと組み合わせて分析に使用する
 - ▶ 形態論情報や韻律情報 x 行為・行為の連なり

13

談話行為タグの出現分布からみる CEJCコアの特徴

14

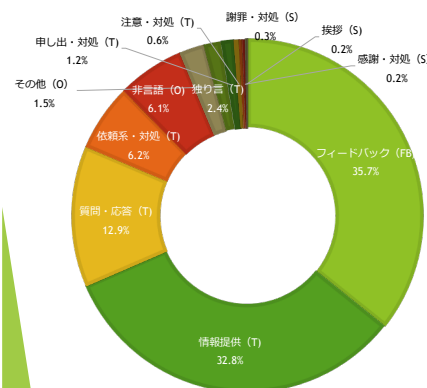
LV1タグ分布：大分類



- ▶ タスク系 + フィードバック系
= 9割
- ▶ その他（多くは「笑い」）
- ▶ 社会的付き合い管理系
= 相対的にみるとわずか
 - ▶ 挨拶は1回の出会いの中で何度も出ない
 - ▶ 謝罪や感謝の表明は文脈が限られる

15

LV1タグ分布：小分類



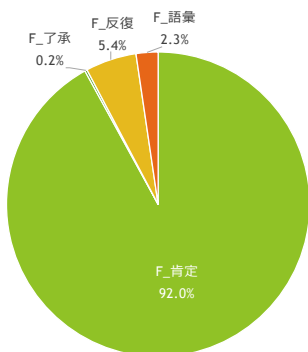
参考) 情報提供 対 フィードバック

- Switchboard Corpus (トピック指定対話) : 53%・20%
- Meeting Recorder DA (合意形成型会話) : 68%・13%

- ▶ 情報提供とフィードバック
= 合わせて7割ほど
 - ▶ 情報提供 対 フィードバック
= 33%・36%
- ▶ 働きかけ - 対処 タイプの発話
 - ▶ 質問・応答
 - ▶ 依頼系・対処
 - ▶ 申し出・対処
 - ▶ 注意獲得・対処
 - = 合わせて2割ほど

16

LV1タグ分布：フィードバック



▶ フィードバック系

- ▶ 肯定
 - ▶ 応答系感動詞 (うん・はい等)
 - ▶ 感情表出系感動詞 (あっ・へー等)
 - ▶ 「そう」系・「そう」系が省略されたもの
- ▶ 了承
 - ▶ オッケー・了解・わかった等 (+文末表現)
- ▶ 反復
 - ▶ 先行発話の一部または全部の繰り返し (+文末表現)
- ▶ 語彙的反応
 - ▶ なるほど・たしかに・本当・わかる等 (+文末表現)

17

談話行為タグと会話の特徴

18

談話行為タグと会話の特徴

- ▶ CEJCに談話行為情報を付与する目的：事例検索
 - ▶ 特定行為やその連なりが生じている箇所をデータから見つけ出すために使用
 - ▶ 会話内の特定部分のやりとりの性質を反映
- ▶ CEJCに談話行為情報を付与する枠組みの特徴
 - ▶ 多様な会話場面に對し共通のアノテーション枠組みを適用
 - ▶ 会話同士の比較が可能
 - ▶ 当該会話全体の性質はどのように反映されるのか？

19

談話行為タグと会話の特徴： やりとりの進み方

- ▶ 比較データ：先生の研究室で、教員と学生の二人会話

▶ T006_001、28分

▶ T010_009、23分

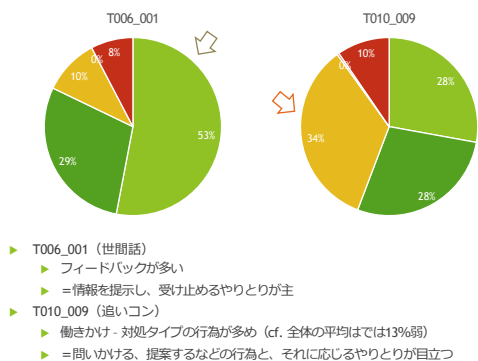


- ▶ 特徴・印象

- ▶ 両データともに親しげ、和やかな雰囲気での対話
- ▶ 話題が異なる (T006_001：スポーツや気候など / T010_009：追いコン)
 - ▶▶ 談話行為にその特徴はどのように反映されているか？

20

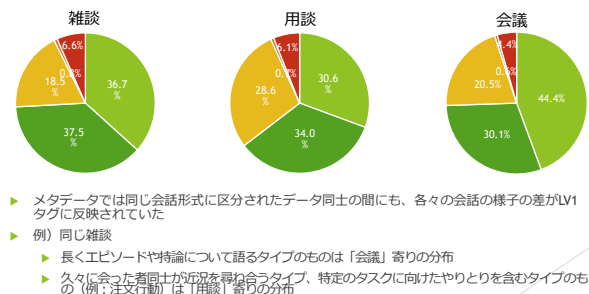
談話行為タグと会話の特徴： やりとりの進み方



21

談話行為タグと会話の特徴： やりとりの進み方

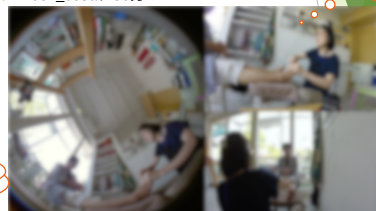
▶ 参考：メタ情報をもとにした会話の形式 x LV1タグ



22

談話行為タグと会話の特徴： 会話のノリ？

- ▶ 比較データ：サービス場面で、サービスを受けながら雑談
▶ T015_008a、12分
▶ K002_003a、30分



特徴・印象

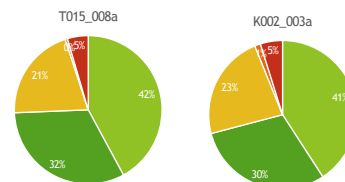
- ▶ 1対1で身の回りのサービスの授受がなされる場面、施術にかかわらない雑談
▶ 作業の間に長めの沈黙が挟まるところなどは似ている
▶ やりとりの箇所は雰囲気はややことなる

まったく

ノリノリ

23

談話行為タグと会話の特徴： 会話のノリ？



LV1タグの分布

- ▶ ほとんど同じ
▶ やりとりの雰囲気に関わるものは、LV1タグでは拾えない情報として生じている

LV2タグの出現数で差が出ているもの

- ▶ 積極的な聞き手反応に関わりそうな振る舞い： 4% > 2.5%
▶ 回収タグ (すでに述べられたこと、予想されることを問い返す振る舞い)
▶ 補完タグ (二人以上で発話を共同構築する振る舞い)
▶ 気安さ、楽しい雰囲気に関わりそうな振る舞い： 0.7% > 0%
▶ 冗談タグ

LV1の生起割合は似ているとしても、各発話の会話の進行において担う役割が異なっていることを示唆

「xxなんだ〜」
「ほんとに？」

24

LV2タグの振る舞い（参考）

- ▶ 全体の2割弱に付与

LV2タグ	出現頻度
付与なし	82.5%
順番取得	3.8%
回収	2.5%
回収への対処	1.8%
順番放棄	1.7%
談話開始	1.6%
談話開始への対処	1.0%

準備系	準備の準備/準備/準備系への対処/準備が対象する本体
回収系	回収/回収への対処/補完/補完への対処
修復系	修復開始/修復操作
談話構造化系	談話開始準備/談話開始/談話開始系への対処/談話終了準備/談話終了/談話終了系への対処
順番管理系	順番取得/順番開放/順番維持/順番放棄
メタ発話情報系	読み上げ/冗談

25

談話行為タグと会話の特徴： LV2タグとの関係

- ▶ 全体の2割弱に付与
- 量的分析にはそれほど向かないが、特定のタグの出現率は会話の一部特性を反映しているように見える
- ▶ 例）補完タグ（二人以上で発話を共同構築する振る舞い）
- ▶ 例（K004_011 1511.589）
1. A: オープンのほうもいっ感じで。
 2. B: 熱しました。
 3. A: はい。| 熱されまして。
- ▶ タグの出現特徴
- ▶ 参加人数が多め、かつ活発に議論がなされている場面に出現が目立つ（多いもので2%弱）
 - ▶ 共感を示すだけでなく、理解を早めにすることでやりとりを前にすすめる使われ方

26

談話行為タグと会話の特徴： LV2タグとの関係

会話ID	発話ID	発話者	LV1	依存関係	依存先	LV2	依存関係	依存先	発話
T015_010	821IC02	高畑	T_情報提供			補完対処	予測的	823	メニューは結構ね <u>このメニューは</u> 。(0.375)いろいろ。(0.232)変えてます。
T015_010	823IC01	小川	T_確認要求			補完	適及的	821	<u>変わるんですよ。</u>
K004_013	565IC04	米山	T_依頼系		放棄				だからこれここに <u>今言ったのが</u> 。
K004_013	568IC02	佐和	F_肯定	適及的	565				うん。
K004_013	569IC02	佐和	T_依頼系対処	予測的	566	補完	適及的		<u>565書かなきゃね。</u>
T004_013	231IC01	一ノ宮	T_情報提供						だ(Wカ)から <u>い</u> 。(0.143)でもだ(Wカ)から(0.127)そうゆう否定するんじゃないかと。(0.222)こう(Dカ)(0.324)こう(0.238)あつ(0.13)あなたはそう思うのね(0.224)つてゆうところを(0.323) <u>こっやうてかな</u> い。(0.126)。(0.307) <u>いっは</u> 。(0.145)(G いや)や それはないでしようって。
T004_013	240IC02	松下	T_情報提供		補完	適及的		231	<u>まあ話しは進まないですよ。</u>
T004_013	515IC05	岩崎	T_情報提供						でもね (W デバート)ディバート)ってゆうのはほんととは(0.123)自分 <u>個人的には</u> 。(0.278)違つ。(.)違うんだけど。
T004_013	517IC01	一ノ宮	T_情報提供		補完	適及的		515	<u>関係ないんだよね。</u>

27

おわりに：今後の課題と展望

28

おわりに

- ▶ CEJCに付与している談話行為タグについて紹介
 - ▶ 事例検索のための使用を想定
- ▶ 談話行為情報をつかってほかの何ができるかについて紹介
 - ▶ LV1タグは
 - ▶ 当該会話全体の特徴：実際にはされているやりとりの面から会話の特徴を大づかみにすることが可能
 - ▶ (応用) 会話内部の特徴：談話の展開なども反映されている
 - ▶ LV2タグは
 - ▶ 感覚的な特徴を多少捉えることが可能かもしれない
- ▶ 今後の課題と展望
 - ▶ 形態論情報、発話長などその他の情報と組み合わせることで分析が可能になっていく
 - ▶ 検索システムでどのようにこれらのデータを扱うかなど、具体的な利用に向けて整備

29

参考文献

- ▶ Allen, J. and Core, M. (1997). Draft of DAMSL: Dialogue act markup in several layers. Department of Computer Science, University of Rochester.
- ▶ Allwood, J. (1997). A critical look at speech act theory. In Dahl, O. (ed.) Logic, pragmatics and grammar. Lund: Studentlitteratur. pp. 53-69.
- ▶ Bunt, H. (2009). The DIT++ taxonomy for functional dialogue markup. In Proceedings of the AAMAS 2009 Workshop Towards a standard markup language for embodied dialogue acts (EDAML 2009), pp. 13-23, Budapest.
- ▶ ISO 24617-2 (2012). Language resource management— Semantic annotation framework (SemaF)—Part 2: Dialogue acts.
- ▶ Shriberg, E., Dhillon, R., Bhagat, S., Ang, J. and H. Carvey (2004). The ICSI Meeting Recorder Dialog Act (MRDA) Corpus. In Proceedings of HLT-NAACL SIGDIAL Workshop, Boston.
- ▶ 居間友里子・佐藤晴 (2018) 「二段階発話連鎖アニメーション：意味・語用論と相互行為」シンポジウム「日常会話コーパス」Ⅲ，国立国語研究所。

30