

国立国語研究所学術情報リポジトリ

学習者の日本語の話し方がわかる「BTSJ日本語自然会話コーパス」：日本人の話し方との比較を通して

メタデータ	言語: Japanese 出版者: 公開日: 2021-03-09 キーワード (Ja): キーワード (En): 作成者: 宇佐美, まゆみ, 張, 未未, 高山, 春花 メールアドレス: 所属:
URL	https://doi.org/10.15084/00003193

学習者の日本語の話し方がわかる『BTSJ日本語自然会話コーパス』

—日本人の話し方との比較を通して—



宇佐美まゆみ (国立国語研究所日本語教育研究領域 教授)・張未末 (国立国語研究所 共同研究員)・高山春花 (国立国語研究所 共同研究員)

Q.「BTSJ日本語自然会話コーパス」とは？

Ans.自然会話の語用論的分析・対人コミュニケーションの分析に適したコーパス！

『BTSJ日本語自然会話コーパス(2020年版)』

- シナリオのない自発的な自然会話を、母語場面、接触場面の初対面会話、友人同士の会話、教師と学生の論文指導場面等のサブ・グループごとに、年齢や性別を条件統制して収集した会話データをまとめたコーパス。
- 1会話20分程度の会話が、377会話(約92時間)収録されている。
- ◎コア会話群(母語場面、初対面会話、友人同士の雑談)142会話、非コア会話群(母語場面と接触場面の教師と学生の論文指導場面や、電話での依頼場面などの様々な場面の会話)235会話を収録！

世界最大規模！

以下のURLからだれでも利用申し込みができる！
https://ninjal-usamilab.info/lab/btsj_corpus/

Q.どんな特徴がある？

『BTSJ日本語自然会話コーパス』の特徴

文字化作業の効率化 押すだけで自動集計できる

会話条件の記載

周辺言語情報の記載

コーディング(研究者独自の観点から)

データ読み込み	会話条件	入力項目	行挿入	行削除	エディット	マクロの保存	集計	印刷	ヘルプ
<p>会話条件の記載: 会話グループ名: 10.台湾人学習者(上記)と日本人の友人同士雑談(女性) 会話条件(会話の話し相手・会話グループ番号・会話の特長を表すもの): 142-10台湾人学習者(上記)と日本人の友人同士 話者記号の凡例: TPA001: Taiwanese Female (Advanced Learner) JF079: Japanese Female 079</p>									
<p>会話内容の例: 75 69 * JF079 (笑いながら)長いよ。 76 70 * TF A001 長いよ。 77 71 * JF079 2,300年はあるんだよ。だって。 78 72 * TF A001 何で? 79 73 * JF079 (笑いながら)中国の... 80 74 * JF079 (笑いながら)中国は長すぎるから(笑い)。 81 75 * TF A001 長すぎる(笑い)。 82 76 * JF079 長すぎる(笑い)1000年じゃあ?。 83 77 * JF079 だよ。 84 78 * TF A001 笑。 85 79 * TF A001 笑。ちよっと、ちよっと、ちよっと(笑い)。 86 80 * JF079 (笑いながら)長いよ(笑い)。 87 81 * TF A001 (笑いながら)長いよ(笑い)。 88 82 * JF079 (笑いながら)長いよ(笑い)。 89 83 * TF A001 (笑いながら)長いよ(笑い)。 90 84 * TF A001 (笑いながら)長いよ(笑い)。 91 85 * JF079 (笑いながら)長いよ(笑い)。 92 86 * TF A001 (笑いながら)長いよ(笑い)。 93 87 * TF A001 (笑いながら)長いよ(笑い)。</p>									

図1: 基本的な文字化の原則 (BTSJ: Basic Transcription System for Japanese)

- 話者の属性や対話相手との関係等の条件を統制して収集した会話の文字化資料と音声や、フォルダごとにまとめている。
 ◎日本人女性ベース初対面同性同士雑談(日本人、台湾人初級学習者、台湾人超級学習者)【音声付】初対面男女、同性同士雑談(同等、目上)【音声付】etc.
- BTSJ(基本的な文字化の原則)によって、語用論的分析に必須の周辺言語情報(フィラー、あいづち、発話の重なり、割り込み、笑い、沈黙等)を豊富に記載している。
- BTSJで文字化されているので、文脈としての談話の流れを追うことができる。
- 話し方に大きな影響を与える社会的要因(場面、話者の属性(年齢、性別等)、対話相手との関係)についての「データ一覧」がある。
- 各研究者が、自分自身の研究目的に応じてコーディング(形式や機能の分類)を行える。

6 連動しているコーディングの自動集計ツールで、基本的な集計ができる: 『BTSJシステムセット』

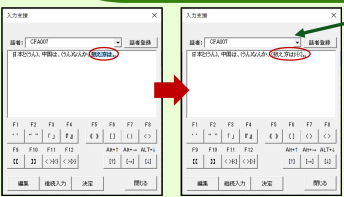


図2: BTSJシステムセット入力支援画面

◎BTSJに基づく文字化の入力支援

◎分析項目の自動集計:

- 総発話文数と話者交替数の算出
- 1会話ごとに、分析項目の頻度と割合を算出
- 複数会話の集計:
 分析項目の頻度と割合の平均/頻度の標準偏差/頻度と割合の最大値・最小値

使い方については動画をご覧ください!

BTSJ活用方法講習会を受ければ入手できる!

https://ninjal-usamilab.info/about_btsj/btsj_seminar/

本研究は、国立国語研究所の機関拠点型基幹研究プロジェクト「日本語学習者のコミュニケーションの多角的解明」、サブ・プロジェクト「日本語学習者の日本語使用の解明」(リーダー: 宇佐美まゆみ)、およびJSPS科研費18H03581「語用論的分析のための日本語1000人自然会話コーパスの構築とその多角的研究」(研究代表者: 宇佐美まゆみ)の成果の一部である。

Q.どんなことがわかる？

Ans.日本語学習者の話し方の特徴がわかる! 相手が初対面であるか、友人であるか、といった会話の状況も考慮できる!

研究例1: 学習者の丁寧体の使い方は？

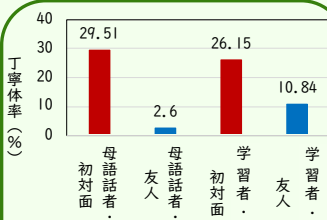


図3: 丁寧体の使用率

- 母語話者、学習者ともに、初対面場面においては、友人場面より丁寧体率が高い。
- 学習者の友人場面での丁寧体率が、母語話者に比べて有意に高い。
- ◎友人と距離を置き、親しくなりにくいというような結果につながるので注意が必要。

友人同士なのに、「です・ます」体で話しています。

◎会話例1:

ライン番号	発話文番号	発話文終了	話者	発話内容
173	160	*	母語話者	じゃ、前京都市に行ったって言ってたよね(はいはい)京都に。
174	161	*	母語話者	その時、京都の八幡を食べた?
175	162	*	学習者	あー、私は日本の食べ物はあー、大好きです(うん)。パンとかビスケットとか(うん)ラーメンとか、焼き鳥、でも、あー、名前は知らないで(うんうん)、私は食べる、何を食べる、これを形を見る時分かりました(あー)、「あー、これこれ」けど、でも、これがいい時、あー、あーこれ名前は何ですか、例えば?
176	163	*	学習者	
177	164	*	母語話者	ピツ、ピツかな。
178	165	*	学習者	あー、ピツ、ピツ食べた?、ピツは何ですか?、(あー)分からない。
179	166	*	学習者	そして、(そっか)どんな、どんなたべものですか?。

研究例2: 学習者は終助詞の「ね」「よ」「よね」を適切に使えているか?

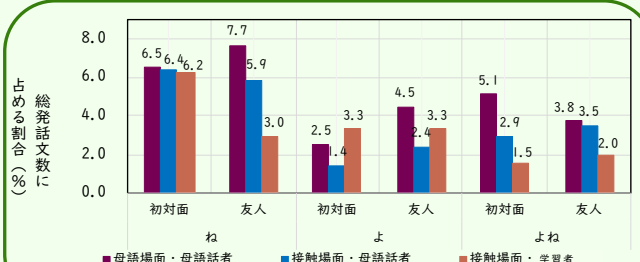


図4: 終助詞「ね」「よ」「よね」の使用率

- 総じて、母語話者のほうが学習者より終助詞の使用率が高い。
- 母語話者は、初対面相手と友人相手による使い分けが明確であるのに対して、学習者は「ね」以外は使い分けが明確ではない。
- 学習者は、使い方によっては押しつけがましくなってしまう「よ」の初対面相手に対する使用率が高いため、日本語教育の観点からは注意が必要。

◎会話例2:

ライン番号	発話文番号	発話文終了	話者	発話内容
226	213	*	学習者	あー、応、(うん)四季があるけど、季節が(うん)、でも、日本のようななんかちゃんとはっきりしてるんじゃない。
227	214	*	母語話者	うんうん。
228	215	*	学習者	冬としても、あたしにとって今のあたしにとってはまあ、(うん)あんまり、寒くないって感じて(うんうんうん)。
229	216	*	学習者	前、春休みの時、向こう帰った(うん)。
230	217	*	学習者	さー、みんなーぼんど、長袖で(うん)、セーターているよ、人(うん)。
231	218	*	学習者	私は半袖で、出掛けた(うん)。
232	219	*	学習者	私「へ?」(笑い)、なんだよ、あの人「うん」ってか、皆がすごい「さー」って(うんうんうん)、友達なんか「TFADII[TFADII]の名前」ちゃん、ちょっとどうしたの? って(うんうんうん)、「夏?」って(笑い)、「夏?」と言われたよ、私(うん)。

Q.何をめざす？

Ans 1. 発話の重なりやあいづち、笑いなど、相互作用に重要な意味を持つ言語現象を定性的分析のみならず、定量的分析も可能にすることによって、語用論的知見の妥当性と信頼性を分野全体として向上させる。

Ans 2. 学習者の日本語の話し方についての分析を、定性的・定量的、両面から深化させ、日本語教育・言語教育の分野に貢献する。

Ans 3. 対話知能学分野の対話システムの研究にBTSJ自然会話コーパスを活用してもらうことによって、対話システム構築に貢献する。