『広島大学日本語電話会話コーパス』：設計と現状

国立国語研究所論集 (NINJAL Research Papers) 22: 113–123 (2022)
ISSN 2186-1358

Corpus of Japanese Telephone Conversation at Hiroshima University: Design and Current Status

IGARASHI Yosuke[a] HIROKAWA Junko

[a]Language Variation Division, Research Department, NINJAL

Abstract
The Corpus of Japanese Telephone Conversation at Hiroshima University (COTCO-H) is a large-scale speech database that is currently under development. COTCO-H contains spontaneous telephone conversations in two different registers (conversations with a local friend and with a campus friend) produced by 50 Hiroshima University students who are native speakers of nonstandard varieties of Japanese. The corpus consists of speech signals and transcriptions for approximately 110,000 words (22 hours), along with morphological annotations such as parts of speech and conjugations. Segmental labeling is currently in progress. COTCO-H also contains different types of read speech produced by the same speakers as auxiliary data. The corpus will contribute to a community of researchers interested in variations across different regions, speech styles, and spontaneity.*

Keywords: Speech corpus, Japanese dialect, spontaneous speech, speech style, sociophonetics

1. Introduction

Several Japanese spontaneous speech databases have been developed over the past two decades, including the Corpus of Spontaneous Japanese (Maekawa 2003), the Chiba Three-party Conversation Corpus (Den and Enomoto 2007), the Corpus of Everyday Japanese Conversation (Koiso et al. 2018), and the RIKEN Mother–Infant Conversation Corpus (Mazuka et al. 2006; Igarashi et al. 2013). The first three are publicly accessible and have been utilized by a wide range of researchers. However, there are few corpus-based resources available for researchers interested in regional variations of Japanese, as these corpora are centered on the standard variety.

The recently released Corpus of Japanese Dialects (Carlino et al. 2019) is the first large-scale speech database of Japanese dialects. Since the release of 24 hours of data in May 2019, more data have been added and released sequentially, and as of February 2021, a total of 60 hours of data had been released (COJADS 2021). Nevertheless, the speech was recorded with a 40-year-old technique, resulting in low-quality recordings, and the utterances of speakers frequently overlapped, which is not ideally suited to acoustic–phonetic analysis. To facilitate

* Many thanks to the audience at NINJAL Salon (April 2021) for feedback on this paper. We thank Hanae Koiso for her valuable advice in the early stages of the corpus construction. We also express our gratitude to Hideaki Kikuchi and Ken'ya Nishikawa for their advice on the method of segmental labeling and to James Tanner for useful comments. This work was supported by JSPS KAKENHI Grant Numbers 17H02332, 19H00530, 16H01933, and 16H03421, as well as the NINJAL projects "Cross-linguistic Studies of Japanese Prosody and Grammar" (project leader: Haruo Kubozono), "Endangered Languages and Dialects in Japan" (project leader: Nobuko Kibe), and "A Multifaceted Study of Spoken Language Using a Large-scale Corpus of Everyday Japanese Conversation" (project leader: Hanae Koiso).

research on regional and stylistic variations in the Japanese language (Takada 2008; Takano and Ota 2017; Tanner et al. 2020), it would be beneficial to construct a high-fidelity recorded speech database that contains various regional dialects produced with different styles and registers.

This paper describes the design and current status of the Corpus of Japanese Telephone Conversation at Hiroshima University (COTCO-H [kɒt.koʊ eɪʧ]), which is currently under construction. COTCO-H is a large-scale speech database that contains spontaneous telephone conversations in various regional dialects of Japanese spoken by 50 students at Hiroshima University. It also contains different types of read speech produced by the same speakers as the auxiliary data. Although two speakers—a participant and an interlocutor—took part in each telephone conversation, the interlocutor's speech was not included in COTCO-H. Thus, COTCO-H cannot be used for conversational analysis, but it does allow the analysis of spontaneous speech produced during a telephone conversation. The corpus consists of high-fidelity recorded speech signals and transcriptions, along with morphological annotations such as parts of speech and conjugations. The size of the corpus of telephone conversations is approximately 110,000 words (22 hours).

The corpus was originally developed by the second author (Hirokawa 2014) and is being extended by the first author at the National Institute for Japanese Language and Linguistics (NINJAL). In the fourth period of the NINJAL Collaborative Research Projects (April 2022–March 2028), the first author will conduct research on the diversity of prosodic systems in Japanese and Ryukyuan dialects, and the construction and analysis of COTCO-H will form part of this research. One reason for publishing the design of the corpus while it is under development is to share it with the potential collaborators who are participating in this research project before its launch.

COTCO-H originally aimed to investigate how students at Hiroshima University modify their speech registers at the university campus (Hirokawa 2014). Although Hiroshima University is located in Hiroshima Prefecture, students from Hiroshima only represent less than 30% of all students (Hiroshima University 2012). The other students are from various (mostly western) parts of Japan, including the Kansai, Chugoku, Shikoku, and Kyushu Districts. Therefore, the native dialects of most students are Western (non-standard) varieties of Japanese, and the students are thus expected to speak differently when they talk to friends at the university and when they talk to their local friends; otherwise, certain aspects of the vocabulary or grammar of their native dialects would not be understood by others.

The corpus contains two types of registers produced by the same speaker: a conversation with a local friend and a conversation with a friend at the university. This design enables a comparison of two registers of the same speaker. In addition to the 22 hours of spontaneous speech, the corpus also contains read speech produced by the same speaker, which allows analysis of the variations in Japanese based on different regions, speech styles, and spontaneity.

2. Design

2.1 Participants

Fifty students aged 18–24 years (mean age 19.7, *SD* ± 1.5) at Hiroshima University participated in the recordings of COTCO-H. They were mostly from the western parts of Japan, namely, Kansai, Chugoku, Shikoku, and Kyushu (including Okinawa) Districts, although three participants were from the eastern parts of Japan: the Aichi, Yamanashi (Chubu District), and

Fukushima (Tohoku District) Prefectures. The participants' birthplaces within the western parts were balanced as much as possible. Their gender was considerably skewed: 43 subjects were female and 7 were male. Therefore, COTCO-H is not suitable for the analysis of gender-based variations. Age-graded variations cannot be analyzed based on the speakers' ages.

The participants' information is summarized in Appendix 1, including the ages of the participants at the time of the first recording. The hometowns of the participants' parents were also recorded but are not reported in this paper. Because a longitudinal analysis of language change was originally intended, we requested all 50 participants to take part more than once in the recordings on different dates (at intervals of more than six months). However, only nine participants responded to this request. Specifically, participants HG1, HG2, HS8, YM2, YM3, TT1, SA1, and KG1 participated twice, and participant TT2 participated three times. Thus, the cumulative total number of participants was 60, while the number of actual participants was 50.

The participants signed a consent form agreeing that the speech signals and speakers' information such as gender, age, and hometown would be used for academic purposes, and that they were planned to be released to the public.

2.2 Recordings

Recordings were completed between 2012 and 2014 at Hiroshima University. Each participant conversation was recorded in a quiet room on campus using a head-mounted condenser microphone (AKG C520). For telephone conversations, the interlocutor's speech was also recorded using a telephone-recording adapter, although this was not included in COTCO-H to protect their privacy. Therefore, COTCO-H cannot be used for discourse analyses. These audio recordings were saved onto an SD card (16 bits, 44.1 kHz) using a linear PCM solid-state recorder (Marantz PMD 660).

In total, three datasets were produced: (1) telephone conversations, which form the bulk of COTCO-H; (2) phoneme-balanced sentences; and (3) voice onset time (VOT) sentences. The first dataset is a recording of spontaneous speech, whereas the latter two are recordings of controlled, read speech.

2.2.1 Telephone conversations

In telephone conversations, we recorded the speech of each participant in the two types of registers, "Local Speech" (LS) and "Campus Speech" (CS). "LS" denotes the recording of a telephone conversation with a local friend or a relative, that is, a person with whom the participant was acquainted in the place where they were born and raised. Therefore, in LS, the dialects of the two speakers were the same, and they were expected to speak in their local dialects.

Meanwhile, CS was the recording of a telephone conversation with a friend on campus, with whom the participant was acquainted after entering Hiroshima University and who was not exactly from where the speaker was born. Therefore, in CS, the two speakers in the conversation had different native dialects, and the main speaker (the participant) was expected to modify his/her speech accordingly.

In both LS and CS, the participant was asked to make a telephone call to an acquaintance (a local friend or relative for LS and a campus friend for CS) and talk to the interlocutor for approximately 10 min. The topic of conversation was not specified in advance. Approximately 10 min of recording per dyad was made (approximately 20 min for each main speaker). The final

sample consisted of 22 hours (approximately 110,000 words).

Since many participants decided on or changed their interlocutors on the day of recording, it was difficult for the interlocutors to spend time giving detailed speaker information in addition to their conversation time. Therefore, the only information about the interlocutor noted was gender and relationship with the participant (parent, sibling, or friend). In the case of CS, it is expected that the hometown of the interlocutor (which is obviously different from that of the participant) influenced the participant's speech; unfortunately, however, information on the hometown of the interlocutor is not available. FK1 and FS1 chose their mothers as interlocutors in the LS, but their mothers had come from prefectures different from the participants. The mother of FK1, who is from Fukuoka Prefecture, is from Miyazaki Prefecture, and the mother of FS1, who is from Fukushima Prefecture, is from Hyogo Prefecture. The other five participants who chose their mothers as interlocutors (HG6, EH4, SN1, MY2, and OK1) were from the same home prefectures as their mothers.

The nine speakers who participated in the recordings more than once—HG1, HG2, HS8, YM2, YM3, TT1, TT2, SA1, and KG1—usually had different interlocutors for each recording. The original purpose of this multiple participation was to analyze whether the speaking style changed over time; thus, the other variables, such as the interlocutors, should ideally be fixed. However, because of the above-mentioned restrictions on the choice of interlocutors, it was impossible to fix such variables.

Overall, the attributes of the interlocutors were not sufficiently controlled, except for whether the participants had made their acquaintance in their hometown or on campus. This limitation of COTCO-H should be noted when conducting an analysis using the difference between LS and CS as an independent variable. However, this limitation does not significantly impair the usefulness of telephone conversations in COTCO-H. They are extremely useful for studying variations across regions and can be used to study variations based on spontaneity when combined with different types of read speech, as described below.

Appendix 2 summarizes, for each conversation, the date of the recording, the age and grade of the participants at the time of the recording, and the gender of and relationship with the interlocutors. Because of recording accidents, the LS of participant HS1 and the CS of participant YN1 were missing.

2.2.2 Phoneme-balanced sentences

The phoneme-balanced sentences in COTCO-H represent a subset of ATR503 (Sagisaka and Uratani 1992) developed by the Advanced Telecommunications Research Institute International. ATR-503 contains phoneme sequences that are acceptable in Japanese in a well-balanced manner and has been widely used in research on speech recognition and speech synthesis. It is also useful for investigating how phonemes in each dialect are phonetically realized in controlled, read speech. ATR503 consists of 10 sets of 50 sentences (from A to J, with the J set comprising only 13 sentences). COTCO-H contains only set A, which consists of 50 sentences. In the recordings, the participants were asked to read 50 sentences aloud (729 words). The phoneme-balanced sentences allow comparisons between read and spontaneous speech produced by the same speakers.

2.2.3 VOT sentences

The VOT sentences were designed to analyze cross-dialectal differences in VOT in a laboratory

setting. This dataset, which contained 24 bimoraic nouns, was originally developed by Takao (2012) and was exploited in the study of the Ibaraki dialect (Utsugi et al. 2013). The dataset has two parts: Part 1 (20 words) and Part 2 (4 words).

The nouns in Part 1 (Table 1) were selected based on three variables: (1) the voicing of the initial consonant (voiceless versus voiced), (2) the place of articulation of the initial consonant (labial versus dorsal), and (3) the following vowel (/i, e, a, o, u/). Coronal consonants (/t, d/) were not included in the second factor to avoid affricated allophones before /i/ and /u/.

The nouns in Part 2 (Table 2) were designed to analyze the effects of the voicing of the initial consonant of the immediately following mora. The place of articulation of the initial consonants of both moras was always dorsal, and the vowels in both moras were always /o/. Thus, the variables in Part 2 were the voicing of the first and second moras' initial consonants (voiceless versus voiced).

These words were embedded in a carrier sentence: X *tte yomuyo* ('It is pronounced X'), where X is the test noun. The final phrase (which is the predicate meaning 'is pronounced') in the carrier sentence differed across participants, as in *yomunen*, *yomutoyo*, and so on, according to their native dialects. The phrases were chosen by the participants themselves to be natural in their dialects. In the recordings, the participants were asked to read a set of sentences aloud three times, with the order of the sentences randomized each time.

Similar to phoneme-balanced sentences, VOT sentences allow comparisons of VOT between read speech and spontaneous speech produced by the same speakers.

Table 1. Words in the VOT sentence (Part 1)

		/a/	/i/	/u/	/e/	/o/
Labial	Voiceless	/pari/ 'Paris'	/piza/ 'pizza'	/puro/ 'professional'	/peke/ 'cross mark'	/poro/ 'polo'
	Voiced	/bara/ 'rose'	/biri/ 'tail-ender'	/buta/ 'pig'	/bero/ 'tongue'	/boro/ 'rag'
Dorsal	Voiceless	/kari/ 'hunt'	/kiri/ 'fog'	/kuri/ 'chestnut'	/keri/ 'kick'	/kori/ 'stiffness'
	Voiced	/gaku/ 'frame'	/giri/ 'duty'	/guru/ 'accomplice'	/geta/ 'clogs'	/goku/ 'phrase'

Table 2. Words in the VOT sentence (Part 2)

		Second mora's initial consonant	
		Voiceless	Voiced
First mora's initial consonant	Voiceless	/koko/ 'one-by-one'	/kogo/ 'archaic word'
	Voiced	/goko/ 'five pieces'	/gogo/ 'afternoon'

2.3 Transcription texts

Transcriptions were conducted manually by the second author based on the visual inspection of speech waveforms using an annotation tool for recordings, ELAN (Max Planck Institute for Psycholinguistics 2012), and the speech analysis software Praat (Boersma and Weenink 2012). Speech signals were divided into "transcription units" at the locations of pauses longer than

200 ms. These physically defined transcription units were then transcribed orthographically, including word fragments and filled pauses. Nonlinguistic events observed outside transcription units, specifically laughter, coughs, and breaths, were also transcribed using specific tags.

Transcription texts in COTCO-H were positioned as inputs for the automatic morphological analysis. Therefore, there are no strict rules for transcription, except that it should follow Japanese orthography (kanji-kana mixed writing). However, there are many points that need to be corrected, such as the extensive use of punctuation marks to prevent errors in automatic morphological analysis. Before the corpus can be released to the public, specific tagging will be necessary, such as withholding the speakers' personal information, as well as distinguishing word fragments and filled pauses from ordinary words. These will be fixed manually after the morphological annotations are corrected.

2.4 Morphological annotations

We morphologically analyzed transcribed utterances. UniDic, a machine-readable dictionary for morphological analysis (Den et al. 2008) provided a morphological annotation of COTCO-H. The unit for identifying a word in UniDic is based on the short unit word (SUW), which is either a mono-morphemic word or a sequence of two morphemes, and broadly corresponds to headwords in ordinary Japanese dictionaries. UniDic provides various morphological information for each SUW, including its lexeme, part of speech, conjugation, and orthography (for details, see Den et al. 2008).

Because UniDic is designed primarily to analyze the standard variety of Japanese, morphological analysis is not necessarily successful when applied to COTCO-H, which contains nonstandard varieties. For example, the utterance *ii-yot-ta* '(he) was saying' in the Hiroshima dialect, comprising the verb *ii* 'say', the progressive auxiliary *yot*, and the past auxiliary *ta* can be incorrectly analyzed by UniDic as *iiyot-ta* '(he) wooed', comprising the verb *iiyot* 'woo' and the past auxiliary *ta*. This analysis occurs because the progressive auxiliary *yot* (its basic form being *yoru*) does not appear in standard Japanese. Similar incorrect outputs of UniDic were manually corrected by the first and second authors.

2.5 Segmental labeling

Segmental labeling is currently in progress. Segmental labels—labels of vowels and consonants—will be provided for the materials in the telephone conversations and phoneme-balanced sentences.

Segmental labels will be provided based on the same scheme developed for the Corpus of Spontaneous Japanese (Kikuchi et al. 2003). Labels are phonemic, and some phonetic events, such as the devoicing of vowels and timing of closure release in stops, are also annotated. We generated most of the segmental labels from phonetic transcription in the morphological annotations described in Section 2.4. We are currently working on automatically aligning the generated segmental labels to the time axis using the monophone hidden Markov model distributed in Julius (Lee et al. 2001)—an open-source Japanese speech recognition engine. These automatically generated segment labels are used as first approximations and subjected to subsequent manual correction. Manual adjustment of segmental labels in 22 hours of data would require considerable human and financial resources. We plan to complete the manual correction of at least half (11 hours) of the data by 2027.

3. Status of the corpus

We recorded the utterances of the 50 participants (the cumulative number of participants was 60) and their transcription and morphological annotation. The total number of words in the corpus is approximately 11,000, and the total time is approximately 22 hours.

The next task is segmental labeling, and automatic labeling using a speech recognition algorithm is ongoing. However, during the process of generating segmental labels, several errors were found in the transcription and morphological annotation. Thus, it is necessary to find and correct these errors before manually correcting segmental labels.

A speech corpus of Japanese dialects with segmental labels has thus far been unavailable. Because segmental labels are essential for sociophonetic analysis, they must be implemented even at a high cost of financial and human resources. During the creation of these labels, we will simultaneously validate and evaluate the corpus and conduct a linguistic analysis together with researchers in corpus linguistics and quantitative linguistics.

Prosodic labeling of COTCO-H would advance research on prosodic systems in Japanese dialects, which the first author will conduct as one of the fourth-period NINJAL Collaborative Research Projects (see Section 1). However, comprehensive prosodic labeling based on the X-JToBI scheme (Maekawa et al. 2002) is currently impossible because no prosodic labeling scheme that can describe the intonation of Japanese dialects has been proposed. Still, it is possible to provide prosodic labels for COTCO-H based on the simplified X-JToBI (Igarashi 2015), which describes only prosodic phrasing (break indices) and boundary pitch movements (BPM). More practically, we may apply a model that only describes BPM, which would allow us to study differences in BPM across dialects and, by comparing the results with those in the Corpus of Japanese Dialects (Carlino et al. 2019), the differences across generations.

4. Conclusion

This paper describes the design and current status of the Corpus of Japanese Telephone Conversation at Hiroshima University (COTCO-H), which is currently under development. We plan to release this corpus to a wider community. For this, we must resolve privacy and human rights-related issues, which include measures of withholding the speakers' personal information (discussed in Section 2.3). To the best of our knowledge, a corpus of Japanese dialects with high-quality speech signals and detailed morphological and segmental annotations has not yet been constructed. Therefore, COTCO-H will make a significant contribution to research investigating variations in the Japanese language.

References

Boersma, Paul and David Weenink (2012) Praat: Doing phonetics by computer [Computer program]. http://www.praat.org/ (Retrieved on April 19, 2012).

Carlino, Salvatore, Yoshimi Yoshikawa and Kazuki Aokayma (2019) Corpus of Japanese Dialects (COJADS). Poster presented at the 3rd International Symposium on Linguistics Patterns in Spontaneous Speech Institute of Linguistics. Academia Sinica, Taiwan.

COJADS (2021) Corpus of Japanese Dialects (COJADS). https://www2.ninjal.ac.jp/cojads/index.html (Retrieved on July 16, 2021).

Den, Yasuharu, Junpei Nakamura, Toshinobu Ogiso and Hideki Ogura (2008) A proper approach to Japanese morphological analysis: Dictionary, model, and evaluation. *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC 2008), 1019–1024.

Den, Yasuharu and Mika Enomoto (2007) A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation. In: Toyoaki Nishida (ed.) Conversational informatics: An engineering approach, 307–330. Hoboken: John Wiley & Sons.

Hirokawa, Junko (2014) Hōgen ni taiō shita keitaiso kaiseki jisho no kakuchō: Hirodai Denwa Kaiwa Kōpasu kōchiku ni sai shite [Extension of the morphological analysis dictionary for dialects: On the construction of the Hiroshima University Corpus of Telephone Conversation.] *Nidaba* 43: 31–39.

Hiroshima University (2012) Applicants by prefecture of their high schools etc. https://www.hiroshima-u.ac.jp/system/files/3852/03_todoufukenn.pdf (Retrieved on January 29, 2021).

Igarashi, Yosuke (2015) Inritsu jōhō [Prosodic information.] In: Hanae Koiso (ed.) Hanashi kotoba kōpasu: Sekkei to kōchiku [Speech corpus: Design and construction,] 81–100. Tokyo: Asakura Publishing.

Igarashi, Yosuke, Ken'ya Nishikawa, Kuniyoshi Tanaka and Reiko Mazuka (2013) Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *Journal of the Acoustical Society of America* 134(2): 1283–1294.

Kikuchi, Hideaki, Kikuo Maekawa, Yosuke Igarashi, Kiyoko Yoneyama and Masako Fujimono (2003) Nihongo hanashikotoba kōpasu no onsei raberingu [Phonetic labeling of the Corpus of Spontaneous Japanese.] *Journal of the Phonetic Society of Japan* 7(3): 16–26.

Koiso, Hanae, Yasuharu Den, Yuriko Iseki, Wakako Kashino, Yoshiko Kawabata, Ken'ya Nishikawa, Yayoi Tanaka and Yasuyuki Usuda (2018). Construction of the Corpus of Everyday Japanese Conversation: An interim report. *Proceedings of the LREC 2018*, 4259–4264.

Lee, Akinobu, Tatsuya Kawahara and Kiyohiro Shikano (2001) Julius—An open source real-time large vocabulary recognition engine. *Proceedings of the EUROSPEECH 2001*, 1691–1694.

Maekawa, Kikuo (2003) Corpus of Spontaneous Japanese: Its design and evaluation. *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition* (SSPR 2003), 7–12.

Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi and Jennifer Venditti (2002) X-JToBI: An extended J-ToBI for spontaneous speech. *Proceedings of the 7th International Conference on Spoken Language Processing* (ICSLP 2002), 1545–1548.

Mazuka, Reiko, Yosuke Igarashi and Ken'ya Nishikawa (2006) Input for learning Japanese: RIKEN Japanese Mother–Infant Conversation Corpus. *IEICE Technical Report* 106(165): 11–15.

Max Planck Institute for Psycholinguistics (2012) ELAN [Computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, the Language Archive. https://archive.mpi.nl/tla/elan (Retrieved on April 19, 2012).

Sagisaka, Yoshinori and Noriyoshi Uratani (1992) ATR onsei dētabēsu [ATR spoken language database.] *Journal of the Acoustical Society of Japan* 48(12): 878–882.

Takada, Mieko (2008) Nihongo gotō yūseion no VOT ni kan suru zenkoku bunpu patan [Geographical pattern of VOT in Japanese initial voiced stops.] *Studies in the Japanese Language* 4(4): 48–62.

Takano, Shoji and Ichiro Ota (2017) A sociophonetic approach to variation in Japanese pitch realizations: Region, age, gender and stylistic parameters. *Asia-Pacific Language Variation* 3(1): 5–40.

Takao, Megumi (2012) VOT no hōgensa ni kan suru kōsatsu: VOT to ta no onsei tokuchō to no sōkan no kanten kara [Discussion of dialectal differences in VOT: In terms of correlation between VOT and other phonetic features.] Unpublished BA thesis, Hiroshima University.

Tanner, James, Morgan Sonderegger and Jane Stuart-Smith (2020) Structured speaker variability in Japanese stops: Relationships within versus across cues to stop voicing. *Journal of the Acoustical Society of America* 148(2): 793–804.

Utsugi, Akira, Kan Sasaki and Yosuke Igarashi (2013) Regional variation of VOT in Ibaraki Japanese. *Proceedings of the 27th annual meeting of the Phonetic Society of Japan*, 119–124.

Appendix 1. Participants' information

	Participant	Birthplace		Birthdate	Age	Gender
		Prefecture	Municipality			
1	HG1	Hyogo	Himeji	March 1994	18	female
2	HG2	Hyogo	Himeji	May 1991	21	female
3	HG3	Hyogo	Kobe	February 1994	18	female
4	HG4	Hyogo	Kobe	December 1991	21	male
5	HG5	Hyogo	Taka	June 1993	18	female
6	HG6	Hyogo	Takasago	September 1994	18	female
7	KT1	Kyoto	Kyoto	February 1995	18	male
8	ME1	Mie	Kameyama	October 1992	19	male
9	SH1	Shiga	Takashima	August 1993	18	female
10	WY1	Wakayama	Gobo	September 1992	20	female
11	WY2	Wakayama	Tanabe	January 1992	21	female
12	EH1	Ehime	Iyo	March 1994	18	female
13	EH2	Ehime	Kamiukena	October 1993	21	female
14	EH3	Ehime	Matsuyama	November 1993	18	female
15	EH4	Ehime	Shikokuchuo	July 1992	22	female
16	EH5	Ehime	Uwajima	September 1993	18	female
17	KW1	Kagawa	Takamatsu	March 1993	19	female
18	TS1	Tokushima	Tokushima	September 1991	20	female
19	HS1	Hiroshima	Fukuyama	January 1994	18	female
20	HS2	Hiroshima	Fukuyama	October 1991	21	female
21	HS3	Hiroshima	Higashihiroshima	January 1995	18	female
22	HS4	Hiroshima	Hiroshima	May 1994	19	female
23	HS5	Hiroshima	Hiroshima	October 1993	19	female
24	HS6	Hiroshima	Kure	December 1993	18	female
25	HS7	Hiroshima	Mihara	August 1991	20	female
26	HS8	Hiroshima	Onomichi	August 1989	22	female
27	HS9	Hiroshima	Sera	November 1992	19	female
28	SN1	Shimane	Masuda	December 1991	20	female
29	YM1	Yamaguchi	Bofu	March 1995	18	female
30	YM2	Yamaguchi	Hikari	July 1993	18	male
31	YM3	Yamaguchi	Shunan	July 1992	19	female
32	YM4	Yamaguchi	Shunan	April 1991	22	female
33	TT1	Tottori	Kurayoshi	November 1990	21	female
34	TT2	Tottori	Tottori	September 1991	20	female
35	MY1	Miyazaki	Hyuga	November 1992	22	female
36	MY2	Miyazaki	Miyazaki	October 1994	18	female
37	FK1	Fukuoka	Fukuoka	September 1990	24	male
38	FK2	Fukuoka	Fukuoka	August 1993	18	male
39	SA1	Saga	Saga	November 1993	18	female
40	NG1	Nagasaki	Goto	November 1994	18	female
41	NG2	Nagasaki	Sasebo	December 1992	20	female
42	KM1	Kumamoto	Kumamoto	December 1992	21	female
43	KM2	Kumamoto	Kumamoto	July 1994	19	male
44	KG1	Kagoshima	Kagoshima	January 1993	20	female
45	KG2	Kagoshima	Kagoshima	November 1990	21	female
46	OK1	Okinawa	Naha	September 1991	21	female
47	OK2	Okinawa	Urasoe	September 1993	20	female
48	AI1	Aichi	Nagoya	November 1992	19	female
49	YN1	Yamanashi	Hokuto	December 1993	19	female
50	FS1	Fukushima	Aizuwakamatsu	July 1991	20	female

Appendix 2. Participants' and interlocutors' information for each conversation

	Participant ID	Age	Grade	LS interlocutor's Gender	Relation	CS interlocutor's Gender	Recording date
1	HG1	18	First year	male	friend	female	Apr. 26, 2012
2	HG1	18	First year	male	friend	female	Nov. 21, 2012
3	HG2	21	Third year	female	friend	female	Apr. 26, 2012
4	HG2	21	Third year	female	friend	female	Nov. 12, 2012
5	HG3	18	First year	female	friend	female	Jun. 8, 2012
6	HG4	21	Fourth year	male	friend	male	Aug. 20, 2013
7	HG5	18	First year	female	friend	female	May 10, 2012
8	HG6	18	First year	female	parent	female	Jun. 10, 2013
9	KT1	18	First year	male	friend	male	Jul. 6, 2013
10	ME1	19	Second year	male	friend	male	May 2, 2012
11	SH1	18	First year	female	friend	female	May 11, 2012
12	WY1	20	Second year	female	friend	female	Nov. 30, 2012
13	WY2	21	Fourth year	female	friend	male	Apr. 23, 2013
14	EH1	18	First year	female	sibling	female	May 13, 2012
15	EH2	21	Third year	female	friend	female	Nov. 27, 2014
16	EH3	18	First year	female	friend	female	May 17, 2012
17	EH4	22	Fourth year	female	parent	female	Nov. 20, 2014
18	EH5	18	First year	female	friend	female	Jun. 8, 2012
19	KW1	19	First year	female	friend	female	Jun. 14, 2012
20	TS1	20	Third year	male	friend	male	May 12, 2012
21	HS1	18	First year	-	-	female	May 15, 2012
22	HS2	21	Fourth year	female	friend	female	Jul. 22, 2013
23	HS3	18	First year	female	friend	female	Jun. 22, 2013
24	HS4	19	First year	female	friend	female	Jun. 13, 2013
25	HS5	19	Second year	female	friend	female	Jun. 6, 2013
26	HS6	18	First year	female	friend	female	May 24, 2012
27	HS7	20	Third year	female	friend	female	May 10, 2012
28	HS8	22	Fourth year	male	sibling	female	Apr. 26, 2012
29	HS8	23	Fourth year	male	sibling	female	Nov. 15, 2012
30	HS9	19	First year	female	friend	female	Apr. 27, 2012
31	SN1	20	Third year	female	parent	female	Apr. 19, 2012
32	YM1	18	First year	female	friend	female	Jul. 11, 2013
33	YM2	18	First year	male	friend	male	Apr. 20, 2012
34	YM2	19	First year	female	friend	male	Nov. 22, 2012
35	YM3	19	Second year	male	friend	female	May 10, 2012
36	YM3	20	Second year	female	friend	female	Nov. 29, 2012
37	YM4	22	Fourth year	female	friend	female	Apr. 30, 2013
38	TT1	21	Fourth year	female	friend	male	Apr. 26, 2012
39	TT1	22	Fourth year	female	friend	male	Nov.19, 2012
40	TT2	20	Third year	female	friend	male	Apr. 27, 2012
41	TT2	21	Third year	male	friend	female	Nov. 15, 2012
42	TT2	21	Fourth year	male	friend	female	Jun. 15, 2013
43	MY1	22	Third year	male	friend	female	Nov. 20, 2014
44	MY2	18	First year	female	parent	female	May 23, 2013
45	FK1	24	Fourth year	female	parent	male	Dec. 12, 2014
46	FK2	18	First year	male	friend	male	May 25, 2012
47	SA1	18	First year	female	friend	female	June 8, 2012
48	SA1	19	First year	female	friend	female	Dec. 5, 2012
49	NG1	18	First year	male	sibling	male	Jun. 8, 2013
50	NG2	20	Third year	female	friend	female	Jul. 24, 2013
51	KM1	21	Fourth year	female	friend	female	Nov. 17, 2014

52	KM2	19	First year	male	friend	male	Jul. 12, 2013
53	KG1	20	Third year	female	friend	female	Jul. 17, 2013
54	KG1	21	Fourth year	female	friend	female	Dec. 8, 2014
55	KG2	21	First year	female	friend	female	May 31, 2012
56	OK1	21	Third year	female	parent	female	Apr. 27, 2012
57	OK2	20	First year	female	friend	female	Jun. 14, 2013
58	AI1	19	First year	male	friend	female	Jun. 8, 2012
59	YN1	19	Second year	female	friend	—	Jun. 10, 2013
60	FS1	20	Third year	female	parent	female	May 17, 2012

『広島大学日本語電話会話コーパス』——設計と現状

五十嵐陽介[a]　　廣川純子
[a]国立国語研究所 研究系 言語変異研究領域

要旨

『広島大学日本語電話会話コーパス』（COTCO-H）は現在開発中の大規模音声データベースである。COTCO-H は，広島大学の日本語非標準変種の母語話者である 50 名の学生が 2 つのレジスター（出身地の友人との会話，キャンパスの友人との会話）で発話した電話会話を格納している。本コーパスには，約 11 万語（22 時間）の音声信号に加えて，その転記および品詞や活用などの形態論情報が付与されている。分節音情報付与作業は現在進行中である。COTCO-H にはさらに補助データとして同じ話者による読み上げ音声も含まれている。COTCO-H は，地域や発話スタイル，自発性などの違いによる言語変異に興味を持つ研究者のコミュニティに貢献するものとなるだろう。

キーワード：音声コーパス，日本語方言，自然発話，発話スタイル，社会音声学