Toward Natural System Repair: An Analysis of Human Other-Initiated Self-Repair Patterns in Japanese Casual Conversations

Kazuya Tsubokura*, Yurie Iribe* and Norihide Kitaoka†
* Aichi Prefectural University, Japan
E-mail: id231001@cis.aichi-pu.ac.jp, iribe@ist.aichi-pu.ac.jp
† Toyohashi University of Technology, Japan
E-mail: kitaoka@tut.jp

***Abstract*—With the advent of large language models (LLMs), dialogue systems have become capable of generating more natural utterances. However, dialogue breakdowns still occur, highlighting the need for effective repair strategies. To enable more natural repair behaviors in dialogue systems, we analyzed the linguistic patterns of repair utterances in human-human communication and investigated how humans perform repair. Based on how speakers signal communication problems, other-initiated repairs (OIRs) have been categorized into three types. Using this classification framework, we analyzed data from the Corpus of Everyday Japanese Conversation and found substantial variations in part-of-speech (POS) patterns across different OIR types. Specifically, we observed differences in the average number of morphemes, the types of POS tags used, and their frequencies for each OIR type. These findings provide fundamental insights that can contribute to the development of dialogue systems capable of performing natural and contextually appropriate repair behaviors.**

I. Introduction

With the advent of large language models, dialogue systems have become capable of generating natural responses. However, just as misunderstandings and mishearings frequently occur in human-human conversations and require repair, it is also essential to detect and appropriately repair breakdowns during human-system dialogues. Various studies have been conducted to detect dialogue breakdowns. For example, the “Dialogue Breakdown Detection Challenge” [1] is a workshop where participants compete in detecting breakdowns in text-based dialogues with users, and numerous methods for breakdown detection have been proposed. On the other hand, although several approaches to repairing dialogue breakdowns have been suggested [2], [3], there is no established method for dialogue repair.

Since repair occurs in human-human communication as well, analyzing how humans perform such repairs can provide valuable insights that can be applied to human-system dialogue. In their study, Ngo et al. [4] identified distinct patterns characterizing other-initiated repair (OIR) utterances for detecting the initiation of repair and generating repair responses in conversational agents. However, their research focuses on task-oriented dialogues, and initiation patterns of repair in non-task-oriented dialogues remain unexplored.

In this study, we focus on “Other-Initiated Self-Repair,” where users initiate a repair following a system breakdown and the system subsequently performs the repair. This study focuses on non-task-oriented Japanese dialogues and analyzes repair behaviors. Building on the classification proposed by Dingemanse and Enfield [5], who identified three types of OIR utterances, we examine the characteristic part-of-speech (POS) patterns associated with each OIR type. The goal of this work is to obtain insights that can inform the implementation of natural repair behavior in dialogue systems. Additionally, assuming that a system can recognize a user’s repair initiated utterance, we conduct a classification experiment to determine whether OIR types can be automatically identified.

The contributions of this study are as follows:

- We reveal part-of-speech patterns for each type of OIR utterance in Japanese casual conversation.
- We perform automatic classification of OIR types and achieve an accuracy of 79.9%.

II. Repair in Conversation Analysis

In conversation analysis, repair is defined as a method for dealing with troubles related to the production, hearing, or understanding of utterances [6]. The process of repair involves the emergence of a trouble source, followed by the marking of the trouble (repair initiation), and then the execution of the repair. For example, consider the following exchange:

A: I forgot that thing. (Trouble Source)

B: What do you mean by “that thing”? (Repair Initiated Utterance)

A: My wallet. (Repair Utterance)

In this case, B does not understand the referent of A’s expression “that thing” and initiates repair by asking for clarification. A then specifies the referent, thus completing the repair.

Repairs can be categorized into four types depending on who initiates and who performs the repair:

- Self-initiated self-repair
- self-initiated other-repair

979-8-3315-7206-8/25/$31.00 ©2025 IEEE

TABLE I
DEFINITIONS AND EXAMPLES OF OTHER-INITIATED REPAIR (OIR) TYPES

OIR Type	Definition and Example
Open Request	A request that signals a problem in the prior utterance without specifying its nature or location. Typically realized using interjections or wh-question words with interrogative intonation. Formulaic expressions like “すみません (Excuse me?)” also fall into this category. Example: “え? (Huh?)”, “何? (What?)”
Restricted Request	A request that narrows down the problem by specifying or characterizing it in more detail, often asking for clarification. Typically realized with question words like “いつ? (When?)” or “どこ? (Which one?)”, and often accompanied by partial repetition. Example (B is the Restricted Request): A: これ持ってる? (Do you have this?) B: どれ? (Which one?)
Restricted Offer	A request that seeks confirmation by specifying the trouble source more precisely. This is done by repeating or paraphrasing all or part of the problematic utterance. Example (B is the Restricted Offer): A: タブレット持ってる? (Do you have a tablet?) B: タブレット? (A tablet?)

- Other-initiated self-repair
- Other-initiated other-repair

In the context of dialogue systems, other-initiated self-repair corresponds to a situation where the system produces an inappropriate or problematic utterance (e.g., a dialogue breakdown), the user initiates repair, and the system then executes the repair. Our ultimate goal is to enable the system to respond appropriately to user-initiated repair attempts. To this end, the present study focuses on other-initiated self-repair, aiming to uncover the linguistic patterns of both repair initiations and subsequent repair utterances. To analyze other-initiated self-repair, we followed the methodology of a prior study that examined the linguistic structure of other-initiated repair (OIR) in task-oriented dialogues [4], and classified each OIR utterance into one of three types as defined by Dingemanse and Enfield [5]. According to Dingemanse and Enfield [5], OIR utterances are categorized into the following three types (see Table I): Open Request, Restricted Request, and Restricted Offer. This classification allows us to investigate linguistic patterns specific to each OIR type. Through this analysis, we seek to provide insights into how a system can recognize the initiation of repair and generate appropriate repair utterances.

III. DATASET

To analyze other-initiated self-repair utterances in Japanese non-task-oriented dialogue, this study utilizes the Corpus of Everyday Japanese Conversation (CEJC) [7]. Repair instances are extracted from CEJC (Section III-B) and annotated with their corresponding OIR types (Section III-C).

A. Corpus of Everyday Japanese Conversation (CEJC)

The Corpus of Everyday Japanese Conversation (CEJC) [7] is a collection of naturally occurring spoken interactions recorded in real-life daily situations. It contains a well-balanced variety of dialogues involving diverse speakers across various contexts. For this study, in order to analyze repair phenomena in human-human interactions, we focus on a subset of CEJC known as the Core, which provides discourse function annotations. The Core consists of approximately 20 hours of data across 52 sessions and includes multiple layers of annotation, making it suitable for detailed dialogue analysis.

B. Extraction of Repair Initiated Utterances

In the Core subset of CEJC, each utterance is annotated with discourse act information based on ISO 24617-2 [8][1]. Among these annotations, the tag `RepairInitiation` indicates the initiation of repair. We first aggregated all utterances annotated with the `RepairInitiation` tag. As a result, out of 59,324 utterances in the CEJC Core, 510 utterances were labeled as `RepairInitiation` in the `dialogAct2` column. [2] Utterances labeled with `RepairInitiation` also included `relation2` column entries that specified either Retrospective Dependence or External Retrospective Dependence.[3] These dependency tags describe the relationship between two utterances performing discourse functions, with the second utterance (i.e., the repair initiation) tagged accordingly. Retrospective Dependence refers to a connection where the second utterance is a voluntary response to the discourse function of the first utterance. In contrast, External Retrospective Dependence indicates responses to non-verbal events (e.g., sounds in the environment), which means such utterances lack an explicit verbal trouble source. Therefore, utterances with External Retrospective Dependence were excluded from the analysis. After this filtering, we obtained 508 repair initiated utterances.

Next, we identified the trouble source utterance corresponding to each repair initiated utterance. Since repair initiated utterances with Retrospective Dependence depend on a prior utterance, we considered those referenced utterances as the trouble source. By referring to the utterance IDs specified in the `relation2` column, we identified candidate trouble source utterances. However, twenty instances where multiple

[1]Discourse act annotation manual: https://www2.ninjal.ac.jp/conversation/cejc/doc/dialogAct_manual.pdf

[2]The `dialogAct2` column in the CEJC Core data provides information about discourse progression and dialogue coordination.

[3]The `relation2` column specifies the ID of the utterance on which the `relationType2` annotation depends.

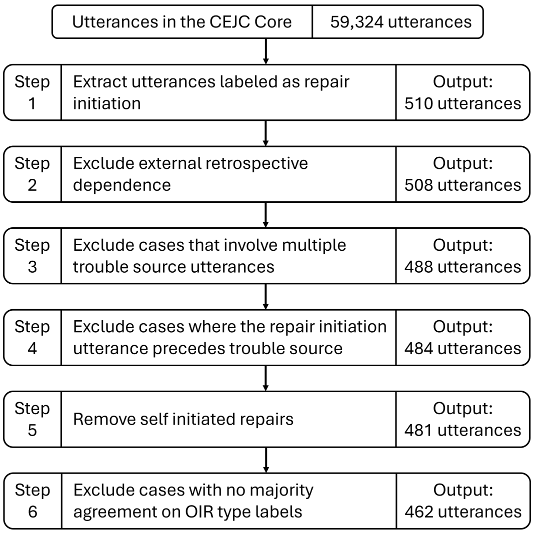

Fig. 1. Flow of extracting other-initiated repair utterances

utterances were indicated as trouble source were also excluded to avoid analytical complexity. Additionally, four instances were excluded because the trouble source was either the same utterance as the repair initiated utterance or occurred later in the conversation sequence. As a result, we obtained 484 repair initiated utterances for further analysis.

Finally, we annotated each of the 484 repair initiated utterances as either self-initiated or other-initiated. If both the trouble source and the repair initiated utterance were produced by the same speaker, it was labeled as self-initiated repair; otherwise, it was labeled as other-initiated repair (OIR). Consequently, only three instances were self-initiated, while 481 instances were classified as OIR. To further identify actual repair utterances, we scanned the CEJC Core for utterances labeled as `Repair` in the `dialogAct2` column. By linking the `relation2` value of each repair utterance to the ID of its corresponding repair initiated utterance, we identified the associated repair utterances. Furthermore, we examined whether the speaker labels of the trouble source and the repair utterance matched, in order to determine whether the repair was a self-repair. As a result, among the 481 OIR utterances, 84 were followed by an actual other-initiated self-repair utterance.

C. Annotation of Other-Initiated Repair (OIR) Types

To analyze other-initiated repair (OIR) utterances, we classified each utterance into one of three types—Open Request, Restricted Request, or Restricted Offer—based on the categorization proposed by Dingemanse and Enfield [5] (see Table I). In this study, we adopted this classification framework as our annotation guideline, and all annotations were performed manually.

Three annotators independently labeled each of the 481 OIR utterances as either Open Request, Restricted Request, Restricted Offer, or Other (e.g., unclear or unclassifiable cases), based on the definitions provided in Table I. Inter-annotator agreement, measured by Fleiss' Kappa, was 0.667, indicating substantial agreement. Final labels were determined based on majority voting among the three annotators. Utterances for which no majority agreement was reached (8 cases) and those labeled as Other by the majority (11 cases) were excluded from the dataset. As a result, 462 utterances were successfully labeled as OIR types, and among them, 79 utterances were followed by a corresponding repair utterance. The procedure for extracting other-initiated repair utterances is shown in the Figure 1.

As a result, the distribution of OIR types was as follows: 249 utterances (53.9%) were classified as Open Request, 149 utterances (32.3%) as Restricted Request, and 64 utterances (13.9%) as Restricted Offer. In contrast, a prior study on task-oriented dialogues [4] reported a distribution of 6.5% Open Request (20 utterances), 10.4% Restricted Request (32 utterances), and 83.1% Restricted Offer (255 utterances). This comparison reveals a clear difference in the distribution of OIR types between task-oriented and non-task-oriented dialogues.

IV. Analysis of Other-Initiated Self-Repair Utterances

A. Method of Analysis

To analyze the linguistic patterns of the three types of other-initiated repair (OIR) utterances, we followed the analytical framework proposed by [4]. We first used GiNZA, an open-source Japanese natural language processing library [9][4], to perform morphological analysis on each repair initiated utterance and extract its sequence of part-of-speech (POS) tags. To examine the relationship between the repair initiated utterance and its corresponding trouble source, we then applied coreference resolution using KWJA [10][5], a general-purpose Japanese language analyzer based on large language models. This allowed us to determine whether the repair initiated utterance referred to the trouble source utterance. Based on the coreference results, we identified the referring words and replaced their corresponding POS tags with the label `[COREF]` in the sequence obtained via GiNZA. We then used this modified POS sequence to analyze grammatical structures by OIR type with Seq2Pat [11][6], a library for extracting sequential patterns. Finally, we analyzed the grammatical structure of the repair utterances using the same procedure as for the repair initiated utterances. Note that transcription tags in the utterance text were manually removed before morphological and coreference analysis.

B. Results and Discussion

1) Distance between the Trouble Source and the Repair Initiated Utterance: As a preliminary analysis, we present a frequency distribution of the utterance distance between the trouble source and the repair initiated utterance. Figure 2 shows that in most cases, the utterance distance is within

[4] https://megagonlabs.github.io/ginza/
[5] https://github.com/ku-nlp/kwja
[6] https://fidelity.github.io/seq2pat/

TABLE II
FREQUENT UTTERANCES AND THEIR FREQUENCIES BY OIR TYPE

Open Request		Restricted Request		Restricted Offer	
Utterance (JP/EN)	Frequency	Utterance (JP/EN)	Frequency	Utterance (JP/EN)	Frequency
ん?。 (Hn?)	84	これ?。 (This?)	11	ランタン?。 (Lantern?)	2
うん?。 (Hmm?)	54	どれ?。 (Which one?)	5	戦争?。 (War?)	2
え?。 (Eh?)	36	何を。 (What?)	4	(Others omitted due to frequency = 1)	
えっ?。 (What!?)	16	どっち?。 (Which?)	4		
何?。 (What?)	8	どうゆうこと?。 (What do you mean?)	3		
何何?。 (What what?)	6	何が。 (What [subject]?)	3		
はい?。 (Sorry?)	6	どこ。 (Where?)	3		
え。 (Uh.)	5	誰?。 (Who?)	3		
あ?。 (Ah?)	4	何が?。 (What [subject]?)	3		
何。 (What.)	4	何を?。 (What [object]?)	2		

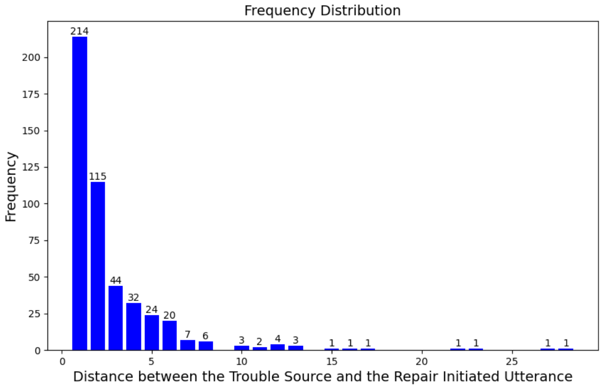

Fig. 2. Frequency distribution of the utterance distance between the trouble source and the repair initiation

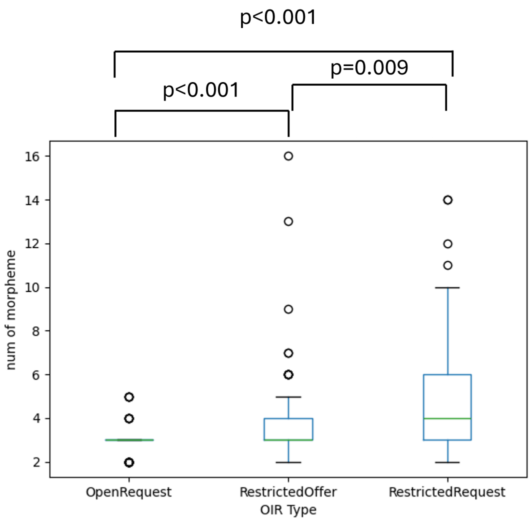

Fig. 3. Number of morphemes in other-initiated utterances by OIR type

five, indicating that repair is typically initiated shortly after the trouble source. This suggests that it is rare for repair to be initiated by referring far back in the conversation, and that focusing on the most recent utterances is effective when identifying the trouble source.

2) Number of Morphemes by OIR Type: We compared the number of morphemes in each of the three OIR utterance types. The number of morphemes in each OIR utterance was calculated using the results of morphological analysis conducted with GiNZA. The average number of morphemes was 2.98 for Open Request, 4.81 for Restricted Request, and 4.16 for Restricted Offer. To test whether these differences were statistically significant, we conducted a Kruskal–Wallis test, which showed a significant difference among the three types ($p < 0.001$). Subsequently, a Steel–Dwass post hoc test confirmed that all pairwise differences were statistically significant at the 1% level, indicating that the number of morphemes significantly varies across OIR types. These results indicate that Open Requests tend to consist of very short utterances, whereas the other two types—Restricted Requests and Restricted Offers—tend to be longer, as they aim to constrain the possible responses.

3) POS Tag Patterns of Repair Initiated Utterance by OIR Type: Next, to analyze the grammatical structure of repair initiated utterances, we examined the POS tag sequence patterns using Seq2Pat. Figures 4, 5, and 6 present the resulting heatmaps for each OIR type. In each figure, the vertical axis indicates the first POS tag in a pair, the horizontal axis indicates the second POS tag, and the values represent the frequency of the corresponding tag sequences. A comparison of the three heatmaps reveals that Open Request utterances involve fewer POS categories, followed by Restricted Offer and then Restricted Request, which shows the most diverse and complex tag patterns. As shown in Figure 4, Open Request frequently includes INTJ (interjections), and many transitions from INTJ to SYM (symbols, such as question marks) were observed—suggesting frequent use of expressions like "ん?。 (Hn?)" or "うん?。 (Hmm?)". In Figure 5, Restricted Request utterances are characterized by a high frequency of ADP (particles), more so than the other OIR types. This aligns with our earlier observation that Restricted Request has the highest average number of morphemes, suggesting that the use of particles contributes to added semantic detail. As shown in Figure 6, Restricted Offer utterances feature a notably higher frequency of COREF (coreference). This indicates that speakers frequently refer back to specific elements in the trouble source utterance via coreferential expressions to signal

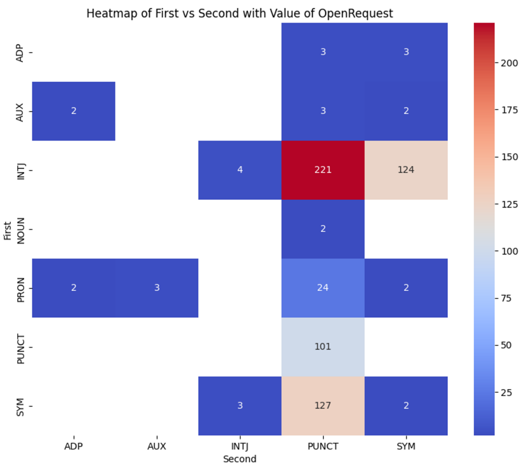

Fig. 4. POS tag sequence patterns for Open Request utterances

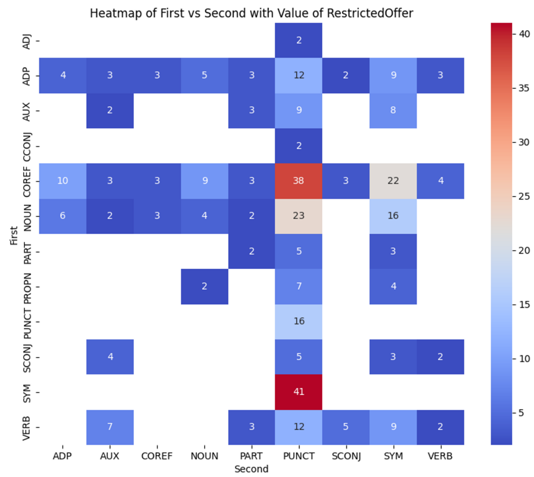

Fig. 6. POS tag sequence patterns for Restricted Offer utterances

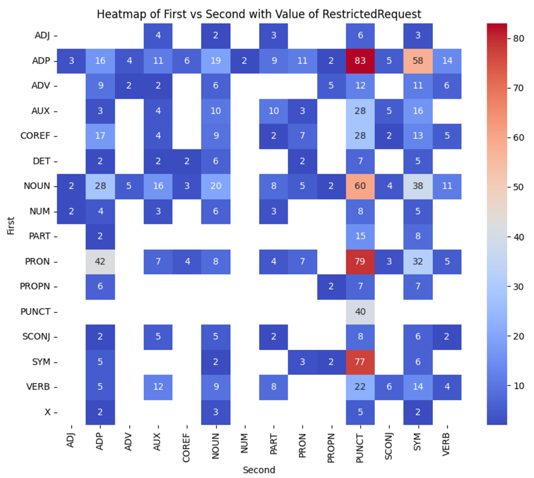

Fig. 5. POS tag sequence patterns for Restricted Request utterances

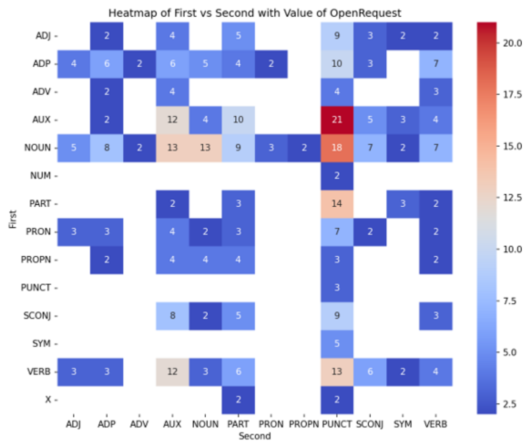

Fig. 7. POS tag sequence patterns for repair utterances after Open Request

misunderstanding or request clarification. In summary, the analysis demonstrates that each OIR type exhibits distinct linguistic patterns.

4) POS Tag Patterns of Repair Utterances by OIR Type: Finally, we analyze the part-of-speech (POS) tag sequence patterns of repair utterances (Figures 7, 8, and 9). A comparison of the three heatmaps reveals that repair utterances in response to Restricted Offers exhibit low POS diversity. This suggests that, since Restricted Offers often initiate repair through confirmation-seeking utterances, the repair can frequently be completed with simple responses such as "うん (yeah)", which require minimal syntactic structure. In contrast, repair utterances following Open Requests and Restricted Requests show greater POS diversity. In the case of Restricted Requests, the system must respond to a specific question or clarification request, and thus the form of repair varies depending on the context. In Open Requests, many responses involve repeating the trouble source utterance or clarifying what was misunderstood, indicating a wide range of repair strategies.

C. Classification of Other-Initiated Repair (OIR) Types

As the previous section confirmed that the grammatical structures differ across OIR utterance types, we next attempted to automatically classify the OIR types. For each repair initiated utterance, we used BERT (`bert-base-uncased`) to extract the embedding vector of the `[CLS]` token. Using these vectors, we performed 5-fold cross-validation with a Support Vector Machine (SVM) classifier using a linear kernel.

The classification achieved an accuracy of 79.9% and an F1-macro score of 0.704. The confusion matrix is shown in Figure 10. From the confusion matrix, we observe that Open Request and Restricted Request utterances were generally well classified, whereas Restricted Offer utterances were frequently misclassified as Restricted Request. These results indicate that repair initiated utterances exhibit distinct linguistic structures depending on the OIR type, and that they can be automatically classified.

V. Conclusions

In this study, we analyzed patterns of other-initiated self-repair utterances in human-human communication within

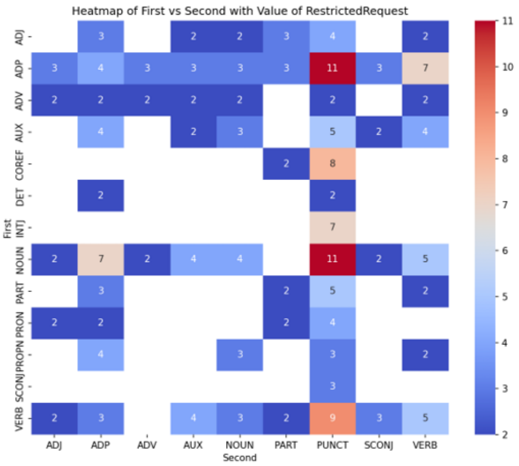

Fig. 8. POS tag sequence patterns for repair utterances after Restricted Request

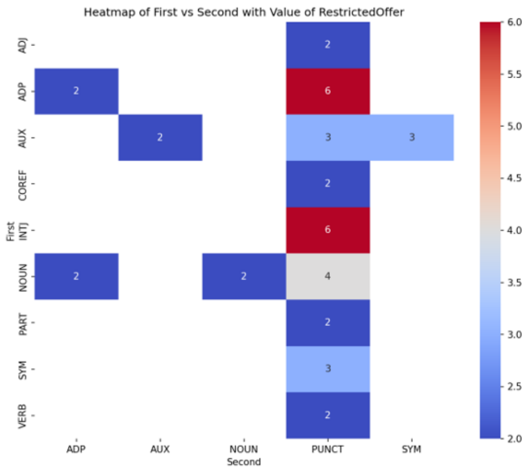

Fig. 9. POS tag sequence patterns for repair utterances after Restricted Offer

Japanese casual conversation, using the Corpus of Everyday Japanese Conversation. We annotated utterances in the corpus with their respective other-initiated repair (OIR) types and examined the part-of-speech (POS) patterns for each type. As a result, we clarified the characteristic POS tag patterns for each OIR type in Japanese dialogue. Furthermore, we conducted automatic classification of OIR utterance types and achieved an accuracy of 79.9%. As future work, we aim to advance the generation of repair utterances and implement repair mechanisms in dialogue systems.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP23H00493.

References

[1] Ryuichiro Higashinaka, Luis F. D'Haro, Bayan Abu Shawar, Rafael E. Banchs, Kotaro Funakoshi, Michimasa Inaba, Yuiko Tsunomori, Tetsuro Takahashi, and João Sedoc. Overview of the Dialogue Breakdown Detection Challenge 4, pp. 403–417. Springer Singapore, Singapore, 2021.

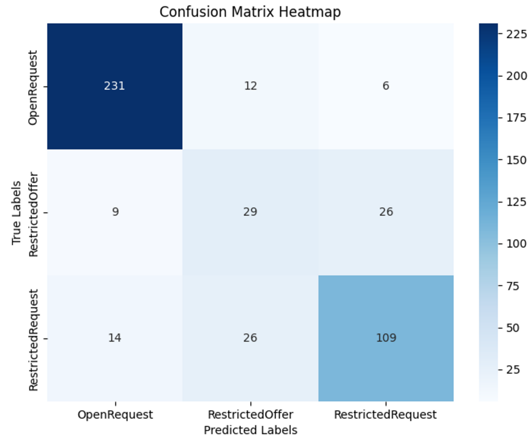

Fig. 10. Classification results for Other-Initiated Repair (OIR) utterance types

[2] Dennis Benner, Edona Elshan, Sofia Schöbel, and Andreas Janson. What do you mean? a review on recovery strategies to overcome conversational breakdowns of conversational agents. In International Conference on Information Systems (ICIS), Dec 2021.
[3] Kazuya Tsubokura, Takuya Takeda, Yurie Iribe, and Norihide Kitaoka. Dialog breakdown recovery strategies based on user personality. In Proceedings of the 14th International Workshop on Spoken Dialogue Systems Technology (IWSDS 2024), 2024.
[4] Anh Ngo, Dirk Heylen, Nicolas Rollet, Catherine Pelachaud, and Chloé Clavel. Exploration of human repair initiation in task-oriented dialogue: A linguistic featurebased approach. In Tatsuya Kawahara, Vera Demberg, Stefan Ultes, Koji Inoue, Shikib Mehri, David Howcroft, and Kazunori Komatani, editors, Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 603–609, Kyoto, Japan, September 2024. Association for Computational Linguistics.
[5] Mark Dingemanse and Nick J. Enfield. Other-initiated repair across languages: towards a typology of conversational structures. Open Linguistics, Vol. 1, No. 1, 2015.
[6] Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks, The Preference for Self-Correction in the Organization of Repair in Conversation. Language, Vol. 53, No. 2, pp. 361-382, 1977.
[7] Hanae Koiso, Haruka Amatani, Yasuharu Den, Yuriko Iseki, Yuichi Ishimoto, Wakako Kashino, Yoshiko Kawabata, Ken'ya Nishikawa, Yayoi Tanaka, Yasuyuki Usuda, and Yuka Watanabe. Design and evaluation of the corpus of everyday Japanese conversation. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, Proceedings of the Thirteenth Language Resources and Evaluation Conference, pp. 5587–5594, Marseille, France, June 2022. European Language Resources Association.
[8] Yuriko Iseki, Keisuke Kadota, and Yasuharu Den. Characteristics of everyday conversation derived from the analysis of dialog act annotation. In 2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (OCOCOSDA), pp. 1–6, 2019.
[9] Hiroshi Matsuda. GiNZA - Practical Japanese NLP based on Universal Dependencies. Journal of Natural Language Processing, Vol. 27, No. 3, pp. 695–701, 2020. （in Japanese）.
[10] Nobuhiro Ueda, Kazumasa Omura, Takashi Kodama, Hirokazu Kiyomaru, Yugo Murawaki, Daisuke Kawahara, and Sadao Kurohashi. Kwja: A unified japanese analyzer based on foundation models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Toronto, Canada, 2023.
[11] Serdar Kadioglu, Xin Wang, Amin Hosseininasab, and Willem-Jan van Hoeve. Seq2pat: Sequence-to-pattern generation to bridge pattern mining with machine learning. AI Magazine, Vol. 44, No. 1, pp. 54–66, 2023.