Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 662–667
Marseille, 11–16 May 2020

Designing Multilingual Interactive Agents using Small Dialogue Corpora

Donghui Lin†, Masayuki Otani*, Ryosuke Okuno† and Toru Ishida‡
†Department of Social Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan
lindh@i.kyoto-u.ac.jp
*Department of Informatics, Kindai University
3-4-1 Kowakae, Higashiosaka City, Osaka 577-8502, Japan
otani@info.kindai.ac.jp
‡School of Creative Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
toru.ishida@aoni.waseda.jp

Abstract

Interactive dialogue agents like smart speakers have become more and more popular in recent years. These agents are being developed on machine learning technologies that use huge amounts of language resources. However, many entities in specialized fields are struggling to develop their own interactive agents due to a lack of language resources such as dialogue corpora, especially when the end users need interactive agents that offer multilingual support. Therefore, we aim at providing a general design framework for multilingual interactive agents in specialized domains that, it is assumed, have small or non-existent dialogue corpora. To achieve our goal, we first integrate and customize external language services for supporting multilingual functions of interactive agents. Then, we realize context-aware dialogue generation under the situation of small corpora. Third, we develop a gradual design process for acquiring dialogue corpora and improving the interactive agents. We implement a multilingual interactive agent in the field of healthcare and conduct experiments to illustrate the effectiveness of the implemented agent.

Keywords: interactive agent, language resource, multilingual service, Wizard of Oz, context-aware, dialogue corpora

1. Introduction

Interactive dialogue agents play essential roles in various tasks. In recent years, major Internet-based companies have been developing popular smart speakers such as Google Home, Apple HomePod, Amazon Alexa and so on. These interactive agents assist users in performing many tasks via voice command. Technologies enabling interactive agents include speech recognition, dialogue generation, and machine learning. A key barrier is that large amounts of data and language resources are needed to develop most key components. Although popular interactive agents provide various general functions, they are not suitable for supporting dialogues in special fields. Unfortunately, entities in many specialized fields find it difficult to develop their own interactive agents due to a lack of language resources such as dialogue corpora. The situation is even more severe when the interactive agents must offer multilingual support, which is now seen as essential in the service industry and public sectors.

To address the above challenge, we aim to develop a general method for designing multilingual interactive agents. We do not intend to realize advanced general smart speakers like those provided by major companies. Rather, we target feasible solutions for various special fields, assuming that such fields have small or non-existent dialogue corpora. To achieve our goal, we need to deal with the research issues raised by the use of small corpora: how interactive agents can support multiple languages, how to realize a dialogue generation mechanism that considers dialogue context, and how to develop a design process for gradual and iterative improvement of the multilingual interactive agents.

In this paper, we propose a framework of designing multilingual interactive agents based on the above three considerations. First, we integrate and customize external multilingual services for interactive agents. These services are provided by Internet-based language service infrastructures like the Language Grid (Ishida et al., 2018) (Lin et al., 2020). The Language Grid can provide necessary language services including specified dictionaries, machine translation, morphological analyzers, etc. In particular, we integrate machine translation and multilingual adjacency pairs for the purpose of multilingual support. Second, we realize a context-aware dialogue generation mechanism for the interactive agents that does not require the existence of large corpus. This mechanism can generate utterances of the agents while considering the context during a dialogue. Third, we propose a process for the gradual acquisition of dialogue corpora via Wizard of Oz (Dahlbäck et al., 1993) (Fraser and Gilbert, 1991), which is a method used for simulating and improving incomplete interactive technologies. In our research, the wizard helps create dialogue manually or selects appropriate dialogue generated by the agents as necessary.

Based on our proposed framework, we implement a multilingual interactive agent for supporting patients in the healthcare field. When a patient speaks to the interactive agent, the agent answers and interacts with him or her continuously. In early stages of the design, utterances of the agent are generated by a hidden wizard who selects appropriate utterances from the candidates output by context-aware dialogue generation module. Moreover, the system accumulates the dialogues between the agent and the end user via Wizard of Oz as new multilingual dialogue examples.

Therefore, the interactive agent can gradually get more accurate dialogue examples until the wizard completes its role. Further, we conduct a case study to illustrate the effectiveness of our work by analyzing both the translation quality and context awareness of multilingual interactive agents.

2. Related Work

In previous research, the architecture of a human-computer dialogue is based on the retrieval model, which consists of a chain of five components: speech recognition, natural language understanding, dialogue manager, natural language generation and speech synthesis (Pietquin, 2005). When a user speaks to the system, the speech recognition accepts the utterance and converts the sound into a sentence. Next, the system retrieves the meaning of the sentence in the process of natural language understanding. The processes, which include speech recognition and natural language understanding, are also called the input system. After the input system, the dialogue manager adopts a strategy that determines how the system should respond to the end user. When the response strategy is found, the natural language generation system makes a sentence and the speech synthesis system converts the words into audio data. The dialogue systems of previous studies can be classified into two major groups: task-oriented systems and chat-oriented systems. The purpose of task-oriented dialogue systems is to help perform certain tasks (Busemann et al., 1997) (Seneff and Polifroni, 2000). On the other hand, chat-oriented dialogue systems not only aim at performing tasks, but also playing games, conducting dialogues or entertaining. For example, IRIS (Informal Response Interactive System) (Banchs and Li, 2012), a typical chat-oriented dialogue system, is an example-based dialogue system based on the vector space retrieval model and implements chat using a dual search strategy from a collection of dialogues in movies. In this paper, we aim to design interactive agents based on the vector space retrieval model for the chat-oriented dialogue system. However, we need to address several issues as we described in the previous section.

Context is recognized as an important consideration in many dialogue generation studies. Williams et. al used a partially observable Markov decision process for a spoken dialogue system (Williams and Young, 2007). The dialogue system is a task-oriented system and supports ticket purchases. Unfortunately, this system cannot satisfy the requirements of chat-oriented systems. In recent years, machine learning approaches have been proposed to deal with dialogue context. For example, Serban et. al used a recurrent neural network (RNN) in order to consider utterance history over a dialogue (Serban et al., 2016). However, most machine learning approaches assume the availability of significant amounts of dialogue corpora. In contrast, we aim to handle the context-aware issue given the existence of only small corpora.

3. Conceptual Framework

To realize the multilingual interactive agents, we need to (1) consider the multilingual support in interactive agents, (2) provide a dialogue generation mechanism that considers dialogue context, and (3) develop a design process for gradual improvement of interactive agents.

3.1. Language Service Integration and Customization for Multilingual Support

In order to solve the first issue, we use external language services including machine translators, morphological analyzers, dictionaries, parallel texts, etc. However, we need to customize the most suitable language services available for different languages and specialized domains. We can use the Language Grid (Ishida et al., 2018) (Lin et al., 2020), a multilingual service platform that allow users to share language resources as services on the Internet. There are two classes of language services provided on the Language Grid: atomic services and composite services. In our proposed framework, we integrate composite machine translation services with domain dictionaries and customize the services for different situations. This last action is extremely important because we assume that the interactive agents will be used in more specialized domains like healthcare, public sectors, shopping centers and so on.

To improve the translation quality for the purpose of realizing enhanced interactive agents, we create a new language service on the Language Grid. The new service uses multilingual dialogue examples including pairs of questions and answers in dialogues in multiple languages, which we call multilingual adjacency pairs, that can well present the dialogue. An example of multilingual adjacency pairs in English and Spanish consist of a set of utterances by two speakers with each utterance in both two languages. One of the pairs has "What is your name?" and "My name is John Smith." The other has "¿Cómo te llamas?" and "Me llamo John Smith." Since the multilingual adjacency pairs are created based on multilingual parallel texts, high translation quality is assured. The multilingual adjacency pair service can be implemented and deployed as a Web service by wrapping the database of multilingual adjacency pairs and providing functions like dialogue search and translation. In our proposed framework, the interactive agents use two language services: the composite machine translation service and the multilingual adjacency pair service. Accumulating more and more multilingual adjacency pairs will improve the translation quality.

3.2. Search System for Context-aware Dialogue Generation

To deal with the second issue, we propose a search system for dialogue recommendation based on the vector space retrieval model. In the search system, the interactive agents make utterances by searching from the multilingual adjacency pairs based on the similarity calculation between the user's latest utterance and the utterances in the adjacency pairs. The system first vectorizes the utterances made in a dialogue between the agent and the user. We use the TF-IDF weighting scheme (Salton and Buckley, 1988), which is the standard bag-of-words weighting scheme and well supports the vector space retrieval model. The vector representation of utterances can be realized based on the work of Han et. al (Han and Karypis, 2000). Then, we use the cosine similarity to calculate the similarity between the utterances.

This basic recommendation method we mention above, however, does not consider contextual information in the dialogue. Therefore, it is not suitable for our purpose. In this paper, we consider the contextual information based on the following two factors. One factor is that the latest utterance by the user does not always provide useful contextual information. For example, "I like a hamburger." is an utterance that does provide contextual information, while "No." is an utterance that does not contribute such information. Therefore, we introduce the concept of *speciality* to represent the degree of an utterance's contribution to the contextual information. The other factor is that the basic recommendation method cannot react when the topic of the dialogue is changed. Hence, we add weights to the vectors when organizing them. Concretely, when appending the weights, it is natural that old utterances are regarded as less important and new utterances are seen as more important. Here we introduce the concept of *freshness* to represent the degree of an utterance's importance to the current context. Then, we modify the basic recommendation method that does not consider any contextual information by combining *speciality* and *freshness* as the weights of each vector of an utterance. Since we focus on introducing the concept of *speciality* and *freshness*, we omit the detailed calculation equations in this paper. It is noted that the proposed method is independent of the size of the corpus, and so can be used with small dialogue corpora.

3.3. Gradual Acquisition of Dialogue Corpora via Wizard of Oz

To realize a context-aware multilingual interactive agent, the third issue is to ensure the quality of the dialogues generated by the agents. However, it is difficult to realize an interactive agent that can provide high quality dialogues automatically from an early stage when the dialogue corpora is small or even does not exist at all. Therefore, it is important to consider how to develop a design process that offers gradual and iterative improvement of the multilingual interactive agents.

Previous studies have widely used Wizard of Oz for prototyping and evaluating dialogue-based computer systems. Furthermore, this method is also applied for designing envisioned systems iteratively. For instance, Klemmer et. al suggested SUEDE (Klemmer et al., 2000), which is a speech interface prototyping tool that provides iterative design. In this research, we apply Wizard of Oz to prototype and design the multilingual interactive agent gradually because the agent has to ensure the quality of dialogues, especially from the viewpoint of translation quality and dialogue breakdown due to missing contexts. Specifically, the system first generates a ranked list of candidate answers by using a search system for dialogue recommendation. The wizard then selects a most appropriate candidate from the list or types a new utterance if there is no appropriate answer in the candidate list. The system uses translation services since we assume that the wizard and the user speak different languages.

To enhance agent quality, our framework enables the system to acquire multilingual adjacency pairs by gradually accumulating new pairs from dialogues. This function is invoked after a wizard selects an appropriate utterance or creates a new utterance to respond to the user. The answer of the wizard and the utterance of the user are regarded as a new adjacency pair. The new adjacency pair is then added to the dialogue corpus together with its translation into every language in the system. The advantages of this function are as follows: (1) since the number of candidates is increased, the range of choices provided to the end user is enhanced; (2) the candidate answers to the questions always include the answer that was actually selected before, i.e., it may be similar to the correct answer as determined by the context. When the system has adequate examples, the agent will be able to select the most appropriate answer greedily with some criteria.

4. Implementation

Fig. 1 shows the implemented architecture of the proposed multilingual interactive agents. The implemented system consists of the following components: (1) the interactive agent interface accessed by end users, (2) the translation service, which translate the user's utterances and the wizard's answers, (3) a search system that receives the translated utterances from the translation service interface and locates, by the context-aware recommendation method, the corresponding candidate answers from the multilingual adjacency pairs, and (4) the interface for the wizard, which shows the candidate answers and receives the answers from the wizard. The database of multilingual adjacency pairs is deployed as the language service to provide accurate translation to the end users.

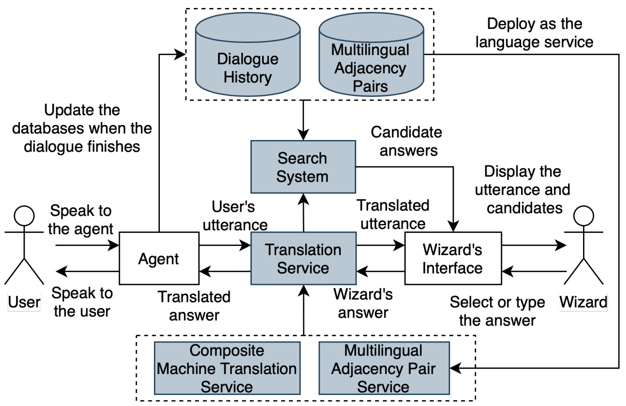

Figure 1: Implementation of multilingual interactive agents

We realize an interactive agent using multilingual dialogue examples from the healthcare field as the initial data. Here we show how the multilingual interactive agent interacts with the user through Wizard of Oz. We explain the case in which the end user is an English speaker and the wizard is a Japanese speaker. First, the end user speaks to the agent. For example, the end user says, "My leg hurts." The system translates the utterance into the wizard's language by the translation service, which can either be a composite machine translation service or a multilingual adjacency pair service. The agent responds to the user through the following process as shown in Fig. 1 : (1) the system searches for candidate answers from among adjacency pairs based on the context-aware method by considering the dialogue history; (2) the search system then provides the candidate answers, which are displayed on the interface to the wizard together with the translated utterance from the user; (3) the wizard selects an answer from among the candidates. For example, the wizard selects "When did the hurting start?" If there is no appropriate answer among the candidate answers, the wizard types a new utterance by free text in Japanese; (4) if the wizard selects an answer from the candidate list, the answer will be translated by the multilingual adjacency pair service. Otherwise, the answer will be translated by the composite machine translation service; (5) the translated answer is delivered to the agent which utters it to the end user via the text-to-speech service; (6) The agent updates the dialogue history and appends new adjacency pairs after speaking to the user. As the dialogue history increases, the performance of the interactive agent is gradually improved because the multilingual adjacency pairs are accumulated during the dialogues.

5. Evaluation

In this section, we conduct experiments in the healthcare field to evaluate the proposed framework. In the experiment, an English speaking user is the patient while the implemented agent is the doctor. We use the criterion of translation quality to confirm the effect of the gradual acquisition of multilingual adjacency pairs, and use the criterion of context awareness to confirm the effect of considering contextual information in improving the interactive agent quality. We do not use the speech recognition service or the speech synthesis service in order to focus on evaluating the dialogue generation mechanism and multilingual support.

The dialogue data used in the experiments was provided by NPO Center for Multicultural Society Kyoto. The data was drawn from scenarios created for developing human resources for medical professionals and translators. This data includes three parallel manual translations: Japanese, English, and Chinese. It has 45 dialogues for hospital situations and 1,206 utterances. The utterances in the data are made by a patient, a doctor and so on. We generated 300 multilingual adjacency pairs from the data.

5.1. Analysis of Translation Quality

We show evaluation results on the improvement of the multilingual interactive agents through the gradual acquisition of adjacency pairs. To prepare for the evaluation, we made 9 blocks from 45 situations from the real data described above. We used 9 time-series phases to simulate the timings of gradual acquisition of adjacency pairs. In the experiment setting, the interactive agent only acquires multilingual adjacency pairs of 5 situations in phase 1. In each next phase, the interactive agent acquires adjacency pairs of 5 more situations. In phase 9, the agent acquires adjacency pairs of all 45 situations from the real data. When the interactive agent acquires more adjacency pairs, there are more chances for the agent to use the multilingual adjacency pair service (higher quality) rather than machine translation service (relatively lower quality) for translating utterances during the dialogues. Therefore, the service quality should increase.

Figure 2 shows the translation quality of utterances in the experiments. Translation quality is evaluated with dialogues of 5 random situations from the real data. The translation quality was scored by 7 bilinguals on a scale of 1 to 5 ("1" is extremely poor, and "5" is perfect). If an ideal multilingual interactive agent were realized, the translation quality would be 5. The horizontal axis in Fig. 2 shows the number of blocks used in the experiment and the vertical one plots the translation quality. The star-marked line indicates the evaluated translation quality achieved when the interactive agent did not use multilingual adjacency pairs and used only machine translation. The value does not depend on the number of corpora and the result of the translation quality is 3.21. The circle-marked line plots the transition in translation quality when the proposed interactive agent was used. When the agent acquired the first block, the result was 3.44. The evaluated translation quality when the agent had all corpora was 4.18. The result shows the effectiveness of the gradual acquisition of dialogue corpora for improving translation quality. The result also implies that the role of wizard is important in the early stages of designing a multilingual interactive agent when only small dialogue corpora are available.

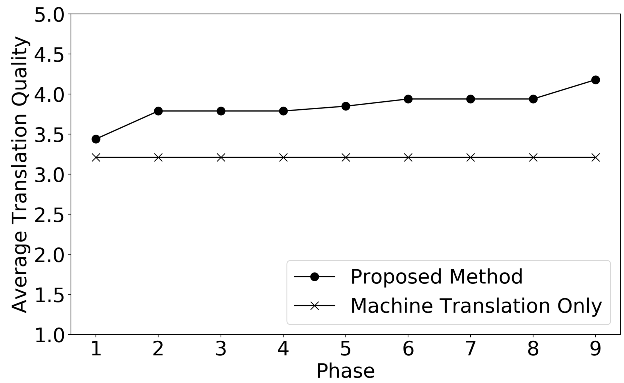

Figure 2: Improvement of translation quality with the gradual acquisition of dialogue corpus

5.2. Analysis of Context Awareness

We first use a case study to show the effectiveness of the search system for context-aware dialogue generation. In the case, the patient has a chest pain and the agent tries to find the cause. We assume the dialogue shown in Table 1 and consider how to recommend the agent to continue this dialogue.

Table 1: Dialogue example between a patient and an agent

Speaker	Utterance
Patient	I have a pain in my chest.
Agent	When did the pain start? (Translated from Japanese)
Patient	When I was running, I felt the pain.
Agent	OK, I see. Do you smoke? (Translated from Japanese)
Patient	No.

We explain the behavior of an agent whose system did not use the contextual information. The system performed a basic cosine similarity calculation between the latest utterance

Table 2: Candidate utterances for the agent in Table 1 without utilizing contextual information

Ranking	Candidate (translated from Japanese)	Score	Appropriateness
No. 1	Do you prefer fatty food such as fried food and meat?	1.0000	Not appropriate
No. 2	Did your parents have convulsions like that in their childhood?	0.4757	Not appropriate
No. 3	Please come back to the examination room again after your tests.	0.4648	Not appropriate
No. 4	I'll wrap tourniquet around your arm and insert a needle. Do you feel numbness in your fingertips?	0.4648	Not appropriate
No. 5	Are you taking any medicine regularly?	0.4648	Not appropriate

Table 3: Candidate utterances for the agent in Table 1 with utilizing contextual information

Ranking	Candidate (translated from Japanese)	Score	Appropriateness
No. 1	Do you still feel pain?	0.6651	Appropriate
No. 2	Do you prefer fatty food such as fried food and meat?	0.3240	Not appropriate
No. 3	Tell me more about the pain. What does it feel like?	0.2700	Appropriate
No. 4	You may be in a condition where your past fatigue emerged all at once.	0.1869	Appropriate
No. 5	Did your parents have convulsions like that in their childhood?	0.1637	Not appropriate

and the utterances in the multilingual adjacency pairs without considering *speciality* or *freshness* described in Sect. 3.2. When the agent received the latest utterance, the dialogue system gave the wizard candidates such as those in Table 2. The scores shown in Table 2 are cosine similarities between the latest utterances of the end user and utterances in the adjacency pairs that have the candidates. The right side of the score indicates whether the candidate is appropriate as a response for continuing the dialogue in Table 1 or not. For example, the adjacency pair of the first candidate is "No." and "Do you prefer fatty food such as fried food and meat?" It means that the latest utterance "No." is identical to the former of the adjacency pair. Thus, the system regards it as the most suitable. However, the candidates are not related to the affected area or the context. It is obvious that such candidates are not suitable as the interactive agent utterances in context. On the other hand, when the system handled the contextual information by considering *speciality* and *freshness*, the system suggested the candidates as shown in Table 3. This case indicates that the proposed search system for context-aware dialogue generation can suggest candidates to the wizard from a broad perspective, especially when a user's utterance has scant contextual information such as the utterance "No."

We also conducted a quantitative analysis by evaluating 8 situations as shown in Table 4. Each dialogue has two exchanges between a patient and the agent. We evaluated candidates which the search system made a suggestion to the wizard when the end user spoke the latest utterance. The latest utterances in situations 1-4 are common phrases and have no contextual information (e.g., "Yes." "Of course." etc.) while those in situations 5-8 are not common and have contextual information (e.g., "I cannot fall asleep at all." "From a week ago." etc.). For each situation, we compare 10 candidates suggested by the two mechanisms: one considers context (the proposed method), and the other does not (the non-context-aware method). We use Mean Average Precision (MAP) of appropriateness of recommended utterances as the evaluation criterion, which is a well-known criterion for information retrieval systems and recommender systems (Sanderson and Zobel, 2005).

Table 4: The 8 situations used for evaluating the context awareness of the proposed method

No.	Situation	Latest Remark
1	Chest pain	No.
2	Breathing problem	Yes.
3	Hyperglycemia	No.
4	Toothache	Of course.
5	Stomachache	I feel pain when I press down.
6	Itchy eyes	My vision is glared.
7	Painful urination	From a week ago.
8	Fever	I cannot fall asleep at all.

Figure 3 and Fig. 4 show the comparison results of MAP values for recommendation appropriateness in 8 different situations. The result in Fig. 3 shows that MAP of the proposed method is higher than that of the method that did not consider the contextual information in situation 1-4. Especially in situation 4, the method that did not use the dialogue history was not able to suggest appropriate candidates and the wizard could not help but input a new utterance and use the machine translation service. The result in Fig. 4 shows that MAP of the proposed method was as high as that of the method that did not consider the contextual information in situation 5-8. In particular, the MAP averages of the two methods are almost the same, which means that the two methods do not differ much if the latest utterance also contains enough contextual information to well guide the recommendation. In summary, the results confirm the effectiveness of the search system for context-aware dialogue generation.

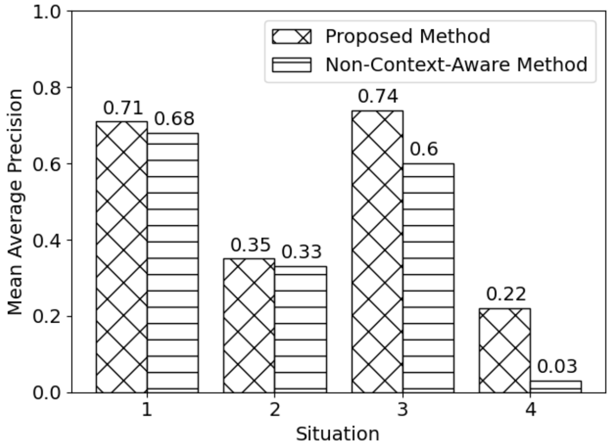

Figure 3: Mean Average Precision of Situation 1-4 where the latest utterances have no contextual information

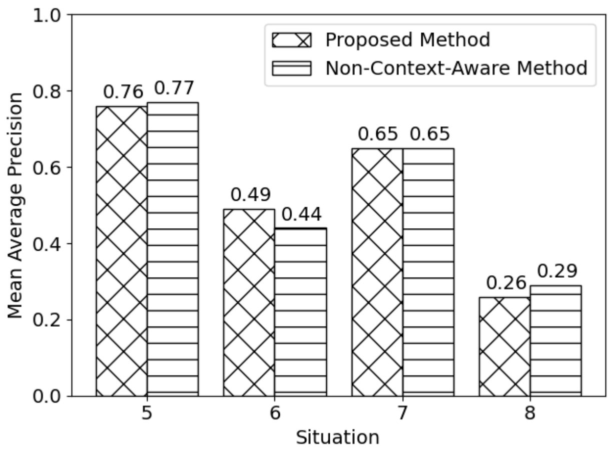

Figure 4: Mean Average Precision of Situation 5-8 where the latest utterances have contextual information

6. Conclusion

The main contribution of this paper is to provide a general framework for designing multilingual interactive agents for various specialized fields wherein only small dialogue corpora are available. Three major components are realized in the proposed framework. First, it permits the integration and customization of external language services for supporting multilingual functions of interactive agents. Second, context-aware dialogue generation is realized in the situation of small corpora. Third, gradual design process is developed for dialogue corpora acquisition. We implemented the proposed framework and conducted experiments to illustrate the effectiveness of the multilingual interactive agents in the aspects of translation quality and context awareness using the data from the field of healthcare. Our future work will mainly focus on improving the search system for context-aware dialogue generation and applying the proposed framework in other specialized fields.

7. Acknowledgements

This research was partially supported by a Grant-in-Aid for Scientific Research (A) (17H00759, 2017-2020), and a Grant-in-Aid for Scientific Research (B) (18H03341, 2018-2020). The authors would like to thank NPO Center for Multicultural Society Kyoto for providing the language resource for multilingual medical dialogues.

8. Bibliographical References

Banchs, R. E. and Li, H. (2012). Iris: a chat-oriented dialogue system based on the vector space model. In Proceedings of the ACL 2012 System Demonstrations, pages 37–42. Association for Computational Linguistics.

Busemann, S., Declerck, T., Diagne, A. K., Dini, L., Klein, J., and Schmeier, S. (1997). Natural language dialogue service for appointment scheduling agents. In Proceedings of the fifth conference on Applied natural language processing, pages 25–32. Association for Computational Linguistics.

Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993). Wizard of oz studies—why and how. *Knowledge-based systems*, 6(4):258–266.

Fraser, N. M. and Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech & Language*, 5(1):81–99.

Han, E.-H. S. and Karypis, G. (2000). Centroid-based document classification: Analysis and experimental results. In European conference on principles of data mining and knowledge discovery, pages 424–431. Springer.

Ishida, T., Murakami, Y., Lin, D., Nakaguchi, T., and Otani, M. (2018). Language service infrastructure on the web: the language grid. *Computer*, 51(6):72–81.

Klemmer, S. R., Sinha, A. K., Chen, J., Landay, J. A., Aboobaker, N., and Wang, A. (2000). Suede: a wizard of oz prototyping tool for speech user interfaces. In Proceedings of the 13th annual ACM symposium on User interface software and technology, pages 1–10. ACM.

Lin, D., Murakami, Y., and Ishida, T. (2020). Towards language service creation and customization for low-resource languages. *Information*, 11(2):67.

Pietquin, O. (2005). A framework for unsupervised learning of dialogue strategies. Presses univ. de Louvain.

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523.

Sanderson, M. and Zobel, J. (2005). Information retrieval system evaluation: effort, sensitivity, and reliability. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pages 162–169.

Seneff, S. and Polifroni, J. (2000). Dialogue management in the mercury flight reservation system. In Proceedings of the 2000 ANLP/NAACL Workshop on Conversational systems-Volume 3, pages 11–16. Association for Computational Linguistics.

Serban, I. V., Sordoni, A., Bengio, Y., Courville, A., and Pineau, J. (2016). Building end-to-end dialogue systems using generative hierarchical neural network models. In Thirtieth AAAI Conference on Artificial Intelligence.

Williams, J. D. and Young, S. (2007). Partially observable markov decision processes for spoken dialog systems. *Computer Speech & Language*, 21(2):393–422.