

Discourse Analysis of the Artificial Intelligence Assistant

Ju-Hyun Lee¹, Tae-Ra Huh² and Hye-Won Han^{3*}

^{1,2,3}Ewha Womans University

¹ththsoso@gmail.com, ²huhtaera313@gmail.com, ³hwhan@ewha.ac.kr

Abstract

The primary purpose of this study is to analyze the components and classify the functions of conversational discourse of the A.I. assistant. Basically, language is a medium for interaction, so to analyze the discourse of conversation can be to analyze of what human and A.I. do with language in a social context. In this study, discourse is viewed as a language use that constitutes and frames a way of talking.

Generally, the addresser (mostly human) and the addressee (mostly artificial intelligence assistant) exchange messages in a common context. In this situation, the contact is offered by a face-to-screen method. According to the modified communication model by structural linguist Roman Jakobson's, types of discourse can be classified by language functions. According to the analysis, the artificial intelligence assistant basically performs the referential and conative functions at this point, and specific functions are prominent, depending on the service type. For example, Siri is an 'empathy-oriented' type that performs the emotive and phatic functions, in addition to the referential and conative functions. Google Assistant is an 'information transfer' type with enhanced referential and conative functions, while Cortana is a 'relation-oriented' type, in which the phatic function overlaps. The speech environment is analyzed through visualization elements strengthened for each service through interface design based on Gestalt theory. It is impossible to perform the metalinguistic function at this stage, but the development of discourse between human and A.I. Assistant will be expanded in various ways through advances in technology. The result of this study can be expected to contribute to a basis for future communication skill in the post-human era.

Keywords: *A.I. assistant, speech recognition, conversational discourse, language function model*

1. Introduction

The era has come when humans can converse with an artificial intelligence assistant. As of 2018, Apple's *Siri*, Amazon's *Alexa*, Google's *Google Assistant*, and Microsoft's *Cortana* are the major and popular artificial intelligence assistants. The A.I. Assistant is mainly an artificial intelligence speech recognition service [1], which can be used through smartphone and speaker. Since the smartphone has the attribute of mobility [2], it enables the user to have a personal conversation at any time at any place.

The A.I. Assistant obtains various kinds of information necessary for communication with humans through machine learning or deep learning. In other words, the source of learning for the A.I. Assistant is human language. It mimics communication between humans, acquires human language, and learns about communication skills. That is, as the dialogue between humans depends on the personality, it is assumed that the assistant's conversation type may also be different.

Table 1 is a list of the representative A.I. Assistants. The A.I. Assistants shown in Table 1 perform conversations according to their own speech principles. Therefore, this study premises that in the dialogue with users, the A.I. Assistants express different persona and subjectivity.

Received (January 5, 2018), Review Result (March 10, 2018), Accepted (March 17, 2018)

* Corresponding Author

Table 1. List of Representative A.I. Assistants (2011-2017)

A.I. Assistant	Developer	Release Year	Speech Principle
Siri	Apple	2011	Accessibility
Watson	IBM	2011	Achieve mutual understanding
Alexa	Amazon	2014	Be brief
Cortana	Microsoft	2014	Natural Interactions
Google Assistant	Google	2015	Grice's Maxims
Houndify	SoundHound	2016	Deep meaning understanding
NUGU	SKT	2016	Who will be anyone NUGU
Bixby	Samsung	2017	SmartThings
Clova	Naver	2017	Compact, Smart, Adorable
Kakao I	Kakao	2017	Connect Everything

Therefore, this study selected the A.I. Assistant mounted with functions that allow speech individualization and discourse generation as the study subjects. Among them, this study chose Apple's *Siri* which established the initial A.I. Assistant structure, Google's *Google Assistant* which is globally used by the public, and *Cortana* which can continuously create the discourse by using the Cortana Skills Kit [3] as the subjects.

The former approaches of the A.I. Assistant are studies in the limited field of technology and marketing [4]. The research on the dialogue model of artificial intelligence has recently begun [5]. But their limitation is that they have regarded the A.I. Assistant as just a kind of service system or application, and limited their roles to accurate and efficient performance. What is important in the case of A.I. Assistant-related studies is that their focus must be placed on what kind of conversations they have with humans, and for what reason, and how the discourses are fundamentally generated.

The purpose of this study is to analyze and classify the components and functions of discourse generation of the A.I. Assistant. Basically, language is a kind of medium for interaction, rather than a system of differences [6]. So to analyze the discourse in the level of talk can be to analyze what human and A.I. do with language in the social context. In this study, discourse, from the theoretical point of view, is conceptualized as a language use, more specifically a set of terms, patterns, and expressions that constitute and frame a way of talking [7].

To prove this, in Chapter 2, major components and functions of discourse between the A.I. Assistant and the user will be suggested, by applying the language model of structural linguist Roman Jakobson. Next in Chapter 3, based on the above extracted discourse model of the A.I. Assistant, types of discourse generation will be classified, and additionally, the user interface will be analyzed.

2. Discourse Components and Functions of the A.I. Assistant

2.1. Discourse Components of the A.I. Assistant

A typical structural linguist, Roman Jakobson, described six universal and common qualities of different languages. According to Jakobson, an *addresser* sends a *message* to an *addressee*, and a *context* and their commonly shared *code* are required for the message to be delivered. Also, the addresser and addressee are physically and psychologically connected through contact to continue communication. His language model was evaluated as a model integrating

the message delivery process with the code concept of sharing meaning [8]. Therefore, Jakobson's model is suitable for an analysis of the discourse of the A.I. Assistant.

Figure 1 shows the language components of Roman Jakobson.

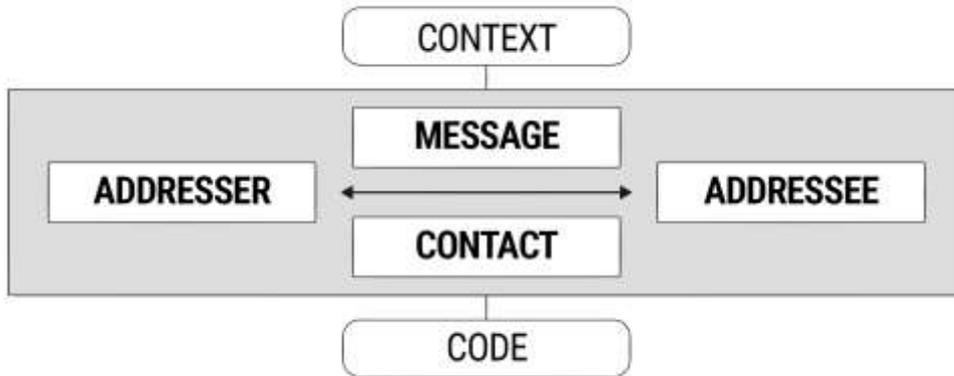


Figure 1. Language Components of Roman Jakobson

The A.I. Assistant basically speaks after learning human language laws and structures. Therefore, the discourse components proposed by Jakobson can be applied to the discourse of the A.I. Assistant, as shown in Figure 2. First, in the A.I. Assistant and human conversational discourse environment, addressers are mostly human, that is, users. Of course, if the technology is more advanced, an A.I. Assistant who can lead the conversation may appear. However, in 2018, it is still limited to the role of the second element addressee. Third, the message between human and A.I. Assistant consists of 'hearable' voice and 'visible' text and image.

With the advancement of multimodality in the future, it is thought that more sensible message expressions will become available, but at the present time, voice and text are the main. Fourth, there is a context surrounding the conversational situation between the human and A.I. Assistant. At present, the context is mainly limited to the sentential context or expressional context. As the A.I. Assistant's technology develops, it is expected that real human beings can consider cultural and personal context that is important in conversation. Fifth, if the contact is basically directed to face-to-face in the conversation between humans, the conversation between human and A.I. Assistant is being implemented in the manner of face-to-screen. At present, however, as the A.I. Assistant has not been humanized, the code, which requires a subjective selection and combination of addressers, is not implemented properly.

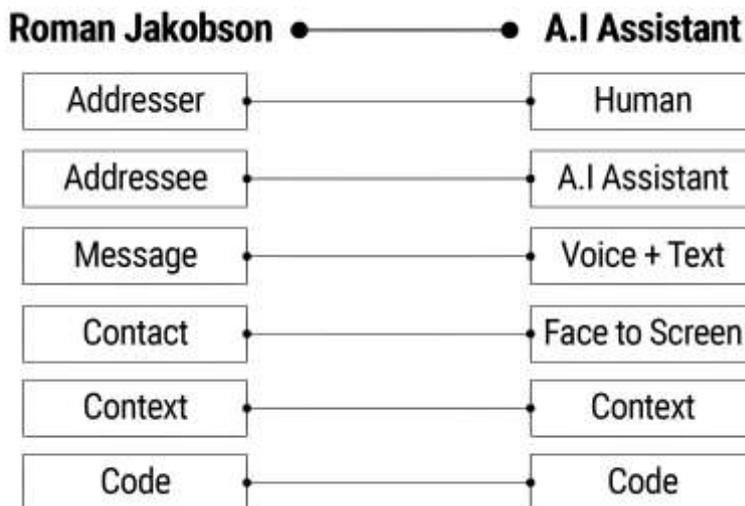


Figure 2. Discourse Components Model of A.I. Assistant

2.2. Language Functions of the A.I. Assistant

The diversity of discourse appears as the components form a hierarchical order. Language functions are divided into five categories, depending on their focus among the components. The first is the referential function that focuses on context; the second is the conative function that focuses on the addressee; the third is the emotion that focuses on the addresser; the fourth is a phatic function that focuses on contact; and the fifth is a metalinguistic function that focuses on the code.

Figure 3 shows the modified Jakobson language function model applied to the A.I. Assistant.

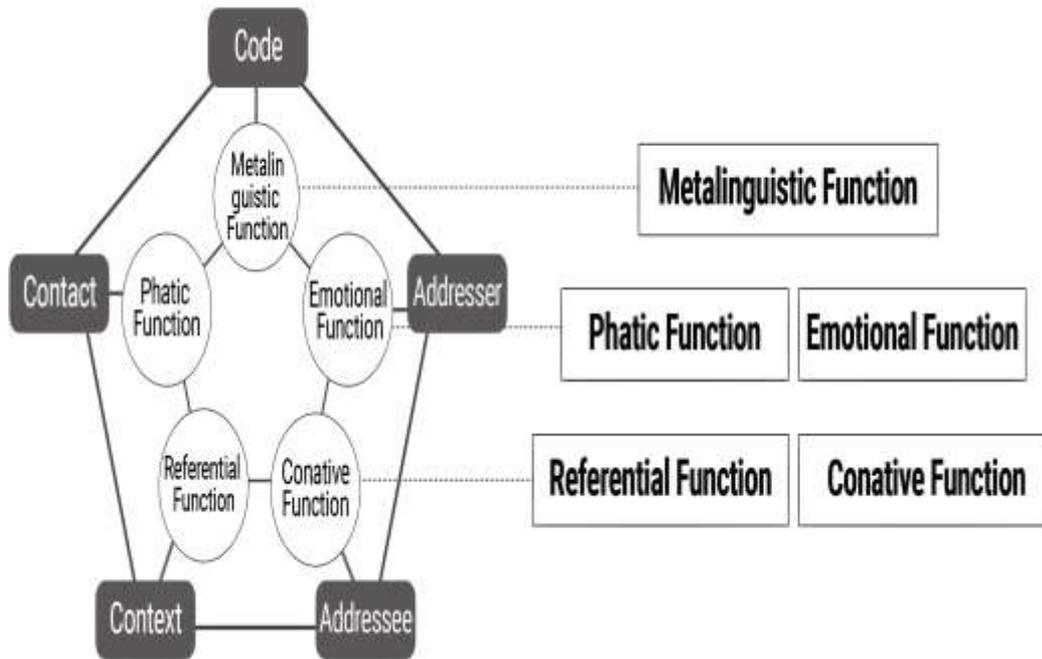


Figure 3. Language Function Model of the A.I. Assistant

As with the discourse of components, Jakobson's language function model is also applicable to the A.I. Assistant. At this time, depending on the A.I. Assistant's conversational principle and the persona being directed, certain functions appear to be enhanced or weakened. First, the referential and conative function is a function that expresses information of empirical reality, aiming mainly at objective and realistic communication. In this case, the function of providing and sharing information is important. The conative function includes those requests, instructions, commands, persuasions, *etc.*, that are made by an addresser to make an addressee believe something, or take some action [9].

The artificial intelligence assistant is basically set up to generate discourse through question and answer. The assistant, which is an addressee, grasps the context of a message sent by the addresser, and presents an appropriate response. For example, it must provide the identical weather information to addressers' various questions that have almost the same meaning, such as "How is today's weather?", "Is it cold today?", and "Do I need an umbrella today?", and at the same time, it needs to consider the context of questions. The study subjects were analyzed to confirm whether the services would be more focused on text than voice in information transfer discourse. The performance of the referential/conative function appeared in all study subjects. The emotive function was added in the case of Cortana. The emotive function and the phatic function additionally appeared in the case of Siri.

Second, the interaction between the addresser and the addressee is enhanced in the emotive and phatic functions. Therefore, it is important to maintain contact and continue dialogue, and

the context of a story is emphasized. The phatic function is the process of establishing and discontinuing communication, and ensuring that communication is taking in place, maintaining the interrelationship and continuing the conversation, ensuring that communication is taking in place [10].

The process of maintaining the relationship and continuing the conversation is important. Therefore, the context of conversation is more significant than the accuracy of the information, and the expression of the emotion and the exchange are important. In the case of the emotive function, the performance of the function is different according to the conversational principle and persona of the A.I. Assistant. For example, if the user says "I want to die", and if the A.I. Assistant displays the telephone number of a nearby consultation center on the screen, it is more like performing a referential/conative function. On the other hand, when it represents answers such as "I am so sorry to hear that", "Please don't try it", and "Please cheer yourself up", it is more like performing the phatic function or the emotive/expressive function.

Finally, the metalinguistic function is a function that can rewrite the language used in the message into its language. This is also a function to check whether the code of the addresser and the addressee exchanges with each other in the communication process are mutually consistent [11]. As a result, since the A.I Assistant at this moment is not recognized as an independent personality, only part of the metalinguistic function can be performed. If an artificial intelligence assistant offers answers similar to "I don't understand it well", or asks another question because it doesn't know how to respond to a request, it performs a kind of metalinguistic function.

The fact that they performed various overlapping functions means that they had the required conditions to create discourse. In addition, the fact they had their unique discourses means that they have a great potential for expansion. Lombard and Ditton have argued that if a media becomes a social actor (medium as social actor), a user can feel its presence [12]. The reason why users feel that they can communicate with the A.I. Assistant is that the technical connection between a machine and voice is preceded by discourses, which can make users feel its presence.

3. Type of Discourse Generation by the A.I. Assistant

3.1. Three Types of Discourse by the A.I. Assistant

Focusing on the language function model in the Chapter 2, if the types of functions performed by the languages of the A.I. Assistant are examined in relation to messages of discourses generated by the services, they can be mainly divided into three types: 'information transfer discourse', 'empathy-oriented discourse', and 'relation-oriented discourse'. In this study, a comparative analysis was conducted by selecting some discourses of the representative A.I. Assistants as samples. Sample discourses contain weather information and how to answer when the user says "I'm bored." As a result, the function and emphasis of discourses varied depending on the A.I. Assistants.

The A.I Assistant subjects have all basically performed the referential/conative function [13]. Therefore, they have provided accurate information about the command or request of the addresser. At this time, the message of the command and the request was interpreted at a level that considers the sentential and expressional contact. However, the information itself was the same, but the way of expressing information in sentences was respectively different according to the A.I. Assistant/s persona, as shown in Table 2.

Table 2. The Discourse Generation Process of the A.I. Assistants

The Discourse Generation Process	Addresser “I’m bored.”		
	↓		
	Addressee		
	Siri	Google Assistant	Cortana
	“I hope it is not because of me.”	“I’ve got some games ready”.	“I can help you with that. Ask me to flip a coin. Heads go out for some fresh air, tails stay in a catch a movie...”
	Addresser “Weather today?”		
	↓		
	Addressee		
Siri	Google Assistant	Cortana	
“Brr, shiver with cold. I think it’s pretty cold.”	“Please check the following.”	“Here’s what the weather’s up to this weekend.”	

First, Google Assistant focused primarily on the referential / conative function, to provide an efficient and concise answer for information delivery. Answers such as “Please check the following.” can be applied to other requests of the addresser, not just to weather information. From this point of view, it can be seen that transferring information itself is a priority. Answers like “I’ve got some games ready.” also have the purpose of transferring information to resolve the user’s situation. The information transfer discourse is the text-dominated one, where the focus is placed on texts, rather than on voice. In other words, if text is made audible, information itself becomes more emphasized, and the referential/conative function plays a greater role. For example, Google Assistant uses the code named ‘Text Balloon’ to make text audible. Google Assistant is the ‘information transfer’ type with enhanced referential and conative functions. That is, the referential / conative function was more emphasized than the relational aspect.

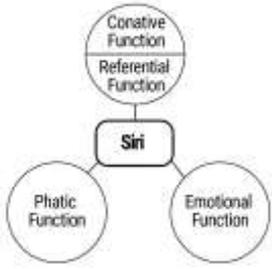
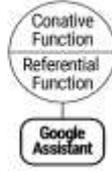
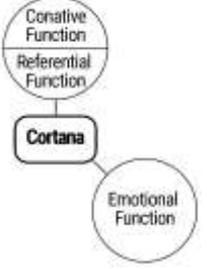
Second, regarding the question for the information similarly, Siri mainly used the discourse markers in the process of answering various kinds of questions through imitative and mimetic words or response tokens such as “Brr”. Expressions such as “shiver with cold” or “pretty” give a sense of naturalness to discourse with the user. Answers like “I hope it is not because of me.” also emphasizes the relationship with the A.I. Assistant. That is, Siri is considered to be able to express empathy while transferring information. From the points of view of expressing its own emotion, and agreeing or sympathizing with the addresser’s message while focusing on the referential / conative function when delivering information, Siri is an empathy-oriented type

that performs the emotive and phatic functions, in addition to the referential and conative functions.

Third, relation-oriented discourse is a type that induces continuity of discourse. In this type, the relationship between the user and A.I Assistant is important. For example, the answer “I can help you with that” means that the relationship between the user and the A.I. Assistant is presupposed. Cortana expresses what information is transferred by directly expressing the words that ‘weather’ while showing the discourse type of enhancing the referential / conative functions like Google Assistant. Cortana also extracts the core content of the information requested by the addresser through self-filtering. It is a differentiating point from Google Assistant, which replies “Please check the following.”

Table 3 shows the three types of discourse generation based on Jakobson’s language function model: Siri, Google Assistant, and Cortana.

Table 3. The Types of Discourse Generation in the A.I. Assistants

	Siri	Google Assistant	Cortana
Types of Discourse			
	Empathy-Oriented	Information Transfer	Relation-Oriented

In addition, the relation-oriented discourse is a voice-dominated one, because voice is visualized; the generation of discourse is more emphasized than information of the discourse. Therefore, the A.I. Assistant’s phatic function to maintain dialogues and relations is more emphasized. In particular, most of A.I. Assistants deliver information in a single turn, while Cortana induces conversation to continue, requesting a following task, such as “Alrighty, what’s next?”, or “Anything else I can do?” Cortana tries to make the conversation persist through the active use of discourse markers, such as ‘next’, and ‘else’, to continue the conversation. In the sense that a discourse can be generated continuously after a single turn, the relationship between the addresser and the addressee, that is, the human and the A.I. Assistant, is strengthened.

Since the A.I. Assistant was based on the assumption that users might request for information, all discourses, regardless of their type, faithfully performed the referential/conative function. In addition, the Q&A system was set up as a default, because the starting point of utterance was a user. Therefore, services can not be classified depending on the presence and absence of information transfer. The differentiated features appeared in the continuation of discourses itself. However, even with the same information, the expression of the answer and the way of ending the conversation differed, depending on whether the A.I. Assistant had an empathy-oriented or relation-oriented persona. This is similar to saying the same message in a unique style by each human personality. Therefore, as the A.I. Assistant persona becomes more complex in the future, the assistant’s conversation is expected to achieve more actuality.

3.2. Interface Characteristics of the A.I. Assistant

The A.I. Assistant basically generates a discourse in the Voice User Interface (VUI). The Voice User Interface is primarily composed of auditory elements, and combines visual elements and tactile elements to implement multimodality in some cases. According to the

interface design based on Gestalt theory, the formative principles were analyzed in relation to visual elements of artificial intelligence assistant interface. Gestalt theory generally transforms information about the form that is transmitted to the cerebrum through vision into a state that can be easily remembered or characterized. It has a tendency to constantly shape external stimuli by giving perceptual behaviors to those who see visual form.

According to the interface design on Gestalt theory of understanding and dealing with human natural visual activities, it is good to see things like proximity, similarity, closed shape, good continuity, and good shape [14]. It urged that there are five types of approaches to interface; the grouping, the structural simplicity, the closure of each element independently, the familiarity that connects the form and interface that the user already knows, the continuity of grouping the visual elements *etc.*

Figure 4 combines the A.I. Assistant's interface approach with the language function model. All three A.I. Assistants, the main subjects, realized five approaches from the perspective of visual perception.

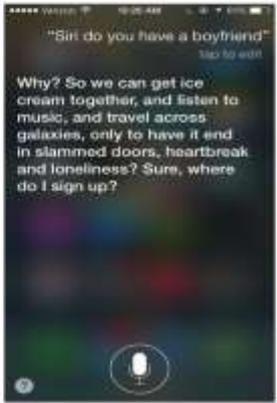
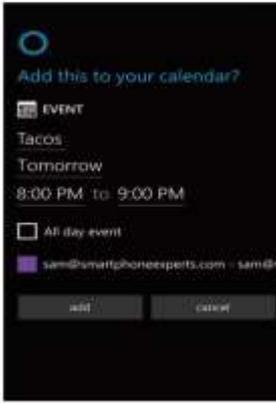
Type	Empathy-Oriented	Information Transfer	Relation-Oriented
A.I Assistant	Siri	Google Assistant	Cortana
Interface Screen			
Interface Element	Grouping	Grouping	Grouping
	Simplicity	Closure	Simplicity
	Familiarity	Continuity	Familiarity

Figure 4. Interface Elements of the A.I. Assistant

First, grouping, the most basic and collective character that gains visibility and readability, is common in all the subjects. In the case of Siri and Cortana, more specifically, grouping (the tendency of various animals, including humans, to form in-groups), simplicity, and familiarity are highlighted. For example, Siri entered the discourse of addresser and addressee as visual text, Cortana summarized the response of addressee and input it as visual text, and maximized efficiency by briefly clarifying the meaning and arrangement of information to convey. In addition, it made puns, and expressed consolation, and sympathy, according to the request of addresser. In the case of Google Assistant, grouping, closure, and continuity were prominent. The elements of obstacles that lead to psychological stability were emphasized, and emphasized continuous elements through the form of exchange of text balloons.

The A.I. Assistant's visual interface sets physical, cognitive, and emotional aspects for optimal user experience. At this time, the visual and tactile elements also work, but in the interface of based on A.I. Assistant, the auditory elements are the main part of the conversation. The user perceives the presence of the A.I. Assistant as a whole through voice communication,

visual interface and tactile reaction. In this respect, the voice interface of artificial intelligence is a 'Formative Interface'. The presence of the addresser is formed through voice conversation with A.I. Assistant, and this existence inherits the possibility that it can be extended to a new relation. With the advancement of technology in the future, the dialogue between A.I. Assistant and human is expected to become more diverse and richer, both in content, and form. This dialogue will be developed as an alternative communication skill in the post-human era.

4. Conclusion

A.I. Assistants form the elements of discourse, and they imitate the language acquisition and communication processes of humans in the process of discourse generation. What is important is the kind of discourses that can be generated between humans and artificial intelligence, and in which forms they are created.

Of course, the discourse of human beings and A.I Assistant at present is still incomplete and becoming increasingly complex. Until now, one-line conversations with the focus on information transfer communication have been mainly developed, but in future, discourses that can express emotions and make relationships with users are expected to be developed. At the current level of technology, the A.I. Assistant cannot perform a completed metalinguistic function, which is possible when a unique persona and language system is completed. At the current stage, the data as input source is interpreted with deep learning, which is utilized for the answer. Therefore, the discourse is mainly focused on functions for information transfer.

However, as the hardware and software supporting the A.I Assistant mature in the future, however, the incomplete language system of the A.I. Assistant will become increasingly perfect. Therefore, a variety of devices and hardware, mainly on the mobile device, will be equipped with the A.I. Assistant, and will naturally form a complex and integrated persona. Therefore, in addition to the function of information transfer, humans and assistants will sympathize and form interactive relationships, and such relationship will be strengthened over time. Humanistic studies of A.I. Assistants are expected to be expanded into post-human studies in the future. Study of the A.I. Assistant's discourse system will enable the A.I. Assistants to have their own unique discourse systems and persona, to strengthen the sense of presence, through active interaction with humans.

It is necessary to search qualitative methodology about how to implement artificial intelligence and what kind of discourse, and ultimately how new communication method of human and non-human in the future post human period.

Acknowledgments

This paper is a revised and expanded version of a paper entitled "Discourse Structure of Artificial Intelligence Speech Recognition Service" presented at the 10th International Conference on Advanced Science and Technology, Ho Chi Minh, Vietnam, February 1-3, 2018.

References

- [1] Y. J. Lee and S. I. Kim, "Study on the Situational satisfaction survey of Smart Phone based on voice recognition technology", *Journal of Digital Convergence*, vol. 15, no. 8, (2017), p. 325.
- [2] J. Ury, H. Lee, "Mobility", Communication Books, Seoul, (2016).
- [3] Y. W. Kim and K. H. Lee, "Microsoft artificial intelligence assistant Cortana", *The Korean Society of Computer and Information*, vol. 25, no. 1, (2017), pp. 8-9.
- [4] Representative studies in the field of technology are as follows.
Y. S. Kim, J. W. Yoon, S. S. Lim, and S. B. Cho, "Implementing a Dialogue Based Emotional Virtual Assistant for Effective Personal Management Services", *Korea Information Science Society*, vol. 27, no. 2C, (2010).
S. J. Yoon, "A Study on New Silver Generation's Emotional Communication and Customized-Virtual Assistant Contents", *Kyunghee Univ. Seoul*, (2017).

H. S. Kim, "Natural Language Processing Techniques of Spoken Assistant Systems based on Artificial Intelligence", Communications of the Korean Institute of Information Scientists and Engineer, vol. 35, no. 8, (2017).

Representative studies in the field of marketing are as follows.

J. M. Kang and J. H. Park, "Research Trend of Artificial Intelligence Service Based on Speech Recognition", The magazine of KIICE, vol. 17, no. 1, (2016).

O. W. Kwon, T. G. Hong, J. X. Huang, and Y. K. Kim, "An Analysis for Dialogue Processing Technologies and Service Trends of Virtual Personal Assistants", Communications of the Korean Institute of Information Scientists and Engineer, vol. 35, no. 8, (2017).

H. T. Yang and D. B. Kim, "Intelligent personal assistant market trend and domestic industry influence forecast", Science and Technology Policy Institute, vol. -, no. 35, (2017).

- [5] H. G. Kim, J. S. Kim, and S. D. Chi, "Design of Character based Conversational AI Using Multi-Agent", The Korea Society for Simulation, vol. -, no. 2017, (2017), p. 11.
- [6] J. Potter and M. Wetherell, "Discourse and Social Psychology: Beyond Attitudes and Behaviour", Sage Publications, London, (1987), p. 146.
- [7] T. J. Watson, "In Search of Management: Culture, Chaos and Control in Managerial Work", Intl Thomson Business Press, New York, (1994), p. 113.
- [8] H. E. Lee, "A Cross-Cultural Study on the Communication Function of Picture Composition in Televised Presidential Debate", Hankuk Univ. of Foreign Studies. Seoul, (2009).
- [9] H. E. Lee, "A Cross-Cultural Study on the Communication Function of Picture Composition in Televised Presidential Debate", Hankuk Univ. of Foreign Studies. Seoul, (2009).
- [10] R. Jakobson, M. S. Shin, "Language in Literature", Moonji publishing, Seoul, (1989).
- [11] H. E. Lee, "A Cross-Cultural Study on the Communication Function of Picture Composition in Televised Presidential Debate", Hankuk Univ. of Foreign Studies. Seoul, (2009).
- [12] M. Lombard and T. Ditton, "At the Heart of It All : The concept of Presence", Journal of Computer-Mediated Communication, vol. 3, no. 2, (1997).
- [13] J. H. Lee, T. R. Huh, and H. W. Han, "Discourse Structure of Artificial Intelligence Speech Recognition Service", Proceedings of the 10th International Conference on Advanced Science and Technology, Ho Chi Minh, Vietnam, (2018) February 1-3.
- [14] A. J. Head, G. S. Park, "Design Wise: A Guide to Evaluating the Interface Design of Information Resources", Gilbut publishing, Seoul, (2000).

Authors



Ju-Hyun Lee,
Doctor's Coure, Department of Content Convergence,
Ewha Womans University, Republic of Korea.



Tae-Ra Huh,
Master's Coure, Department of Content Convergence,
Ewha Womans University, Republic of Korea.



Hye-Won Han,
Associate Professor, Department of Content Convergence,
Ewha Womans University, Republic of Korea.