Challenges and Opportunities in the Spoken Language Understanding in Korean - An NLP perspective

최현정 (HyunJeong Choe), Google Assistant NLP, Senior Staff Linguist

1. Introduction

- **This paper presents current challenges of Korean Spoken language understanding from an NLP(Natural Language Processing) perspective, especially when the NLP is applied for an intelligent virtual assistant.**

Korean Language

NLP

Intelligent Virtual Assistant
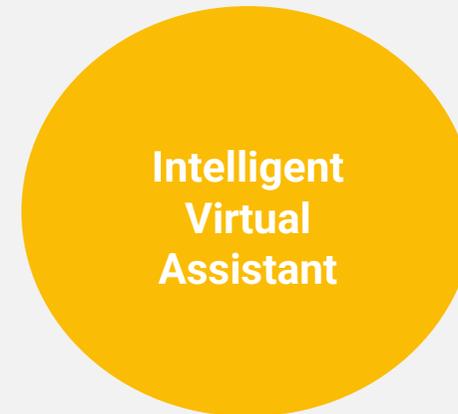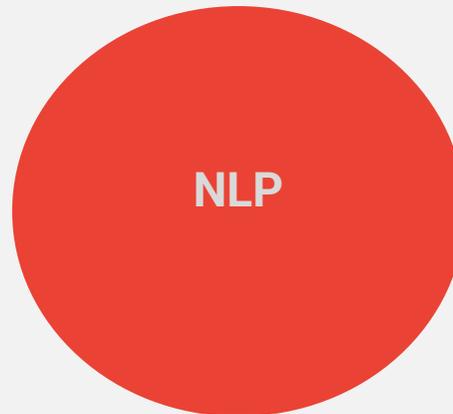
2. Background: Natural Language Processing

- Natural Language Processing (NLP) is a sub-discipline of computer science providing a bridge between natural languages and computers. It helps empower machines to understand, process, and analyze human language. (Amirsina Torfi et al., 2021)

- There are various applications of NLP
   1) Intelligent virtual assistant
   2) Machine Translation
   3) Social Media Monitoring
   4) Survey Analysis
   5) Targeted Advertising etc.,

2. Background: Intelligent Virtual Assistant

- In the last decade the concept of intelligent virtual assistant has become widely known and gained large popularity. (Giancarlo Iannizzotto et al., 2018)

- Several commercial devices such as Naver Clova, Kakao Mini, Google Assistant are able to interact with the user by voice commands in Korean, offer several network-based services and can interface with smart home automation systems, enhancing them with an advanced user interface.

- NLP(Natural Language Processing) is one of core technology of the intelligent virtual assistant.
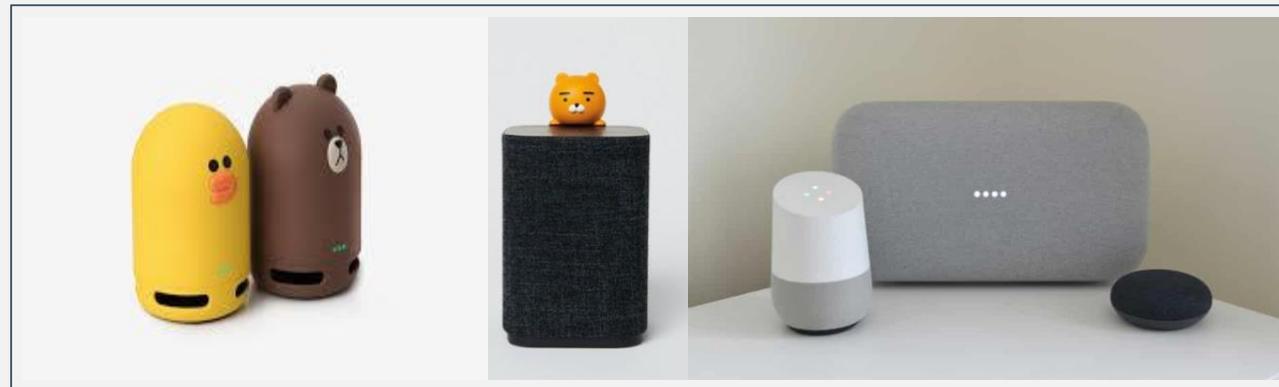
Various commercial intelligent virtual assistant devices

2. Background: Intelligent Virtual Assistant

- "Wake up words" such as Hey Google, or any customized words for each dialog turn was one of big roadblocks to implement "human-human"-like conversational environment.

- Major AI Assistants such as Google and Alexa currently have functionality of continuous conversations without the wake up word each time. (e.g., Google Assistant "Continued Conversation", Alexa "follow-Up mode")

Google Assistant - continuous conversation (bleepingcomputer article, June 21, 2018)

2. Background: Intelligent Virtual Assistant

- There are differences between Human-AI Assistant conversation and Human-Human conversation, especially when they use Korean language. The differences are mainly caused by:

  - **Paralinguistics Communication**: the part of communication outside of the words **themselves** – the volume, speed, intonation of a voice along with gestures and other non-verbal cues. Whenever there is confusion or stereotyping in cross-cultural communication, paralinguistics is most often responsible. (Crystal et al., 1974)

  - **Sociolinguistics -** Language use symbolically represents fundamental dimensions of social behavior and human interaction. The notion is simple, but the ways in which language reflects behavior can often be complex and subtle. Furthermore, the relationship between language and society affects a wide range of encounters--from broadly based international relations to narrowly defined interpersonal relationships. (Wardhaugh, Ronald, 1992)

3. Challenges - Honorifics and Politeness

- Korean language has well-developed **Speech Levels** in terms of politeness.
- The politeness is determined based on multiple complex factors such as relationship between speakers.
- In many conversational scenarios between Human and intelligent virtual assistant, honorifics and politeness play a role to complicate spoken language understanding for the intelligent virtual assistant to perform their task.

[Example]
Human: 김 교수님한테 나 곧 도착한다고 문자 보내. (Text Professor Kim I will be there shortly)

AI Assistant: 네, 김 교수님한테 **"나 곧 도착한다."** 라고 문자를 보냈습니다.
AI Assistant: 네, 김 교수님한테 **"저 곧 도착합니다."** 라고 문자를 보냈습니다.

- Identifying an appropriate speech level and its honorifics is very difficult in the case of the example above.

Honorifics and Politeness [continued]

- Relationship between **human and intelligent virtual assistant** is uncertain, or can't be generalized; requires highly personalized speech styles in Natural Language Generation

  [Example]
  Human: 내일 비 와? (Is it going to rain tomorrow?)

  AI Assistant: 네, 내일 비 예보가 있습니다. 우산을 챙겨 가시기 바랍니다.
  AI Assistant: 네, 내일 비 예보가 있네요, 우산을 챙겨 가시는 게 좋겠어요.
  AI Assistant: 내일 비 예보가 있네, 우산 챙겨 가는 게 좋겠어.
  AI Assistant: 내일 비 온대, 우산 챙겨 가라.

- People may define their relationship with AI Assistant in various ways; as a friend, as a secretary, as a partner etc.
- This affects the speech style of the AI Assistant that the people may want to hear.
- No research or data corpus exist about this problem.

3. Challenges - Pragmatics

- Korean language has very wide range of the **Adjacency Pair** (대화의 인접쌍) depending on **Sociolinguistics factors**.
- Building **Spoken Language corpus** with various of the adjacency pairs (pre-condition of Machine Learning based NLP systems) is hard.

[Example 1]
Speaker 1: 너 원피스 예쁘다. (I like your dress.)
Speaker 2: 아, 이거 싼 거야. (This is cheap.)

Speaker 1: 너 원피스 예쁘다. (I like your dress.)
Speaker 2: 아 진짜? (Really?)

Speaker 1: 너 원피스 예쁘다. (I like your dress.)
Speaker 2: 나도 그렇게 생각해. (I think so too.)

[Example 2]
Speaker 1: 너무 고마워요. (Thank you so much.)
Speaker 2: 아, 아니에요. (No, not at all.)

Speaker 1: 너무 고마워요. (Thank you so much.)
Speaker 2: 제가 더 고마워요. (I more thank to you.)

Speaker 1: 너무 고마워. (Thank you so much.)
Speaker 2: 야 그런 말 하지 마. (Hey do not say so.)

3. Challenges - Pragmatics [continued]

- **Speech Act** (Austin 1962) is an utterance that serves a function in communication. Speech acts are performed when speakers offer an apology, greeting, request, complaint, invitation, compliment or refusal.
- Due to complex Sociolinguistics factors (e.g., politeness level depending on relationship between speakers) in Korean language and lack of using **Paralinguistics** signals (e.g., Prosody, pause) in current NLP, understanding human level of Speech Act is hard for Machine.

      [Example]
      Speaker 1: 뭐 하고 있어? (What are you doing?)
      Speaker 2: 나 계속 기다리고 있었어. (I've been waiting for you.)

- Literal meaning (Semantics): I've been waiting for you.
- Depending on context, this utterance can serve multiple Speech Acts; apology, complaint, invitation.
- Without paralinguistics signals, knowing about the exact context, difficult to understand speech acts.

3. Challenges - Ambiguity

- Korean sentences are often highly ambiguous.

  [Example]
  내일 어디 갈 거야
     I am going to go somewhere tomorrow.
     Are you going somewhere tomorrow?
     Where are you going tomorrow?

- Multiple factors contribute to ambiguity
  - Pro-drop
  - Neutral sentence ending
  - indeterminates (어디)

- Korean speaker can disambiguate this without trouble because of the distinctive prosody of each meaning.

3. Challenges - Ambiguity

- Prosodic disambiguation is hard problem in the current NLP.

[Example 1]
The utterance is ambiguous. Depending on the pitch pattern, its meaning will be different like below:
Human: 그만해. (Stop.)
AI Assistant: 레시피를 중단합니다. (Sure, stop showing the recipe.)
AI Assistant: 아니요, 눌어붙지 않게 계속 저어 주셔야 합니다. (No, you have to continue to stir.)

[Example 2]
The utterance is ambiguous. Depending on the pitch pattern, its speech act will be different like below:
Human: 잘했다. (Good job.)
AI Assistant: 감사합니다. (Thank you.)
AI Assistant: 죄송합니다. (I apology.)

Conclusions

- This paper presents current challenges of Korean Spoken language understanding (in below) from an NLP perspective, especially when the NLP is applied for an intelligent virtual assistant.
  - Ambiguity
  - Honorifics and Politeness
  - Pragmatics

- Although NLP and AI Assistant have been rapidly evolving, there is few **in-depth research in Korean linguistics** that identifying honorifics & politeness, pragmatics factors, and ambiguity handling **in a context of** Human-Computer Interaction (HCI). In-depth linguistic researches of those areas will be very beneficial to advance the NLP for Korean language.

- In addition, building **right linguistic corpora** with various linguistic phenomena is a key for NLP systems. Researches about methodology of building linguistic corpora, particularly with prosodic dimensions of pragmatic phenomena will be useful.

References

- Austin, J. L. (1962). How to do things with words. Cambridge, MA: Harvard University Press
- Giancarlo Iannizzotto et al. (2018). A vision and speech enabled, customizable, virtual assistant for smart environments. Conference: 11th International Conference on Human System Interaction HSI 2018
- Amirsina Torfi, Rouzbeh A. Shirvani, Yaser Keneshloo, Nader Tavaf, Edward A. Fox. (2021). Natural Language Processing Advancements By Deep Learning: A Survey. arXiv:2003.01200v4
- Wardhaugh, Ronald. (1992). *An introduction to sociolinguistics.* Cambridge, MA: Blackwell.
- Crystal, David. (1974) "PARALINGUISTICS: ". *Part 1*, Berlin, Boston: De Gruyter Mouton. pp. 265-296.