A Quantitative Verification of Politeness Theory Using Large Language Models from Japanese Dialogue Data

Tetsuro Takahashi[1]　Mayumi Usami[2]

([1]Kagoshima University　[2]Tokyo University of Foreign Studies)

Background: Dialogue Systems

- Desirable dialogue systems (Engineering)
  - Useful, Accurate, Convenient systems
- Evaluation criteria: propositional accuracy
  - Can it answer questions?
  - Can it execute specific tasks (e.g., ticket booking)?
  - Can it continue the dialogue without breakdowns?
- Meeting these criteria was not easy for a long time

Voice Assistant Systems
- Apple Siri (2011)
- Amazon Alexa (2014)
- Google Assistant (2016)

- **This has become feasible with deep learning language models.**

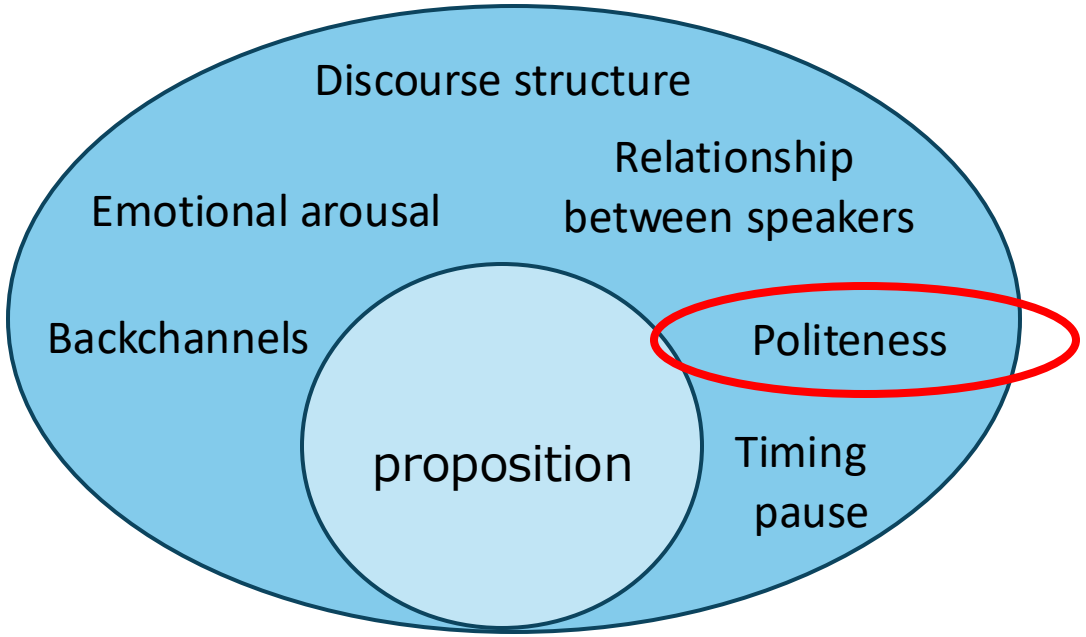
Background: Dialogue Systems

- Newly required perspectives
  - Interpretation of discourse structure
  - Consideration of dialogue context and speaker relationships
  - Emotional changes induced by dialogue
  - Use of paralinguistic cues such as backchannels and timing …

These aspects are not yet sufficiently implemented in engineering-oriented dialogue systems.

Toward the next level of dialogue systems

- Implement insights from dialogue research in linguistics and social sciences.



Research question: ***Can computers handle politeness, not just correctness?***

Politeness Theory

- A framework that systematizes **linguistic strategies** for establishing and maintaining smooth human relationships. (Brown and Levinson 1987)

Humans have two basic needs.

Positive face

The desire to be understood, liked, and appreciated by others.

Negative face

The desire not to be interrupted , imposed upon, or restricted by others.

When people communicate, they often threaten the other person's face.

Politeness is understood as a linguistic strategy used to compensate for face-threatening acts.

Politeness Theory

- Politeness is regarded as proportional to the degree of face-threatening acts.
- Brown and Levinson proposed Equation (1) to calculate the degree of face threat W_x as a basis for estimating politeness.

$$W_x = D(S,H) + P(H,S) + R_x \quad (1)$$

W_x	Degree of face threat, indicating how much an act (x) threatens the listener's face.
D	Social distance between the speaker (S) and the hearer (H).
P	Power of the hearer (H) over the speaker (S).
R_x	Rank of imposition: cultural weight assigned to an act (x) based on the burden it places on the listener.

Objective of this study

- Objective
  - We examine whether large language models (LLMs) can handle Politeness Theory.

- Verification items:
  - Estimate the values of Rx defined in Politeness Theory and the politeness level of utterances (Po), using a LLM.
  - Quantitatively verify correlation between face-threatening degrees (Wx) and politeness levels (Po) using a dialogue corpus.

Data：BTSJ Corpus

Balanced Transcripts of Spoken Japanese Corpus; BTSJ Corpus （Usami 2023）

BTSJ 1000 Person Japanese Natural Conversation Corpus

BTSJ1000人日本語自然会話コーパス

Home

Site opening date : 2023-12-04

The transcripts and audio files of the BTSJ Japanese 1000-Person Natural Conversation Corpus (BTSJ Corpus), managed by Mayumi Usami, are available from this page.

To use them, be sure to read the conditions of use before downloading the transcripts and audio files.

*The BTSJ Corpus does not include videos, which can be viewed on the Natural Conversation Resource Bank (NCRB).

- Regarding the BTSJ Japanese 1000-Person Natural Conversation Corpus (NCRB-linked completed version, March 2023) [jp]
- Basic Transcription System for Japanese (BTSJ) [jp]

Data：BTSJ Corpus

One advantage of BTSJ is that it includes diverse speaker attributes and relationships, which are considered to potentially influence politeness.

Dialog No.	Characteristics of dialogue	# dialogues	Total time
1	Casual Talk Between same-gender friends	19	444 min 24 sec
2	Between friends and first-time acquaintances	23	482 min 5 sec
3	Between teacher and student	10	311 min 00 sec
:	:	:	:

Data：BTSJ Corpus

An example of dialogue (original)

2	会話フォルダ名： 04. 同性同士の断りの電話会話(対先輩、対同級生、対後輩)(女女)【音声付】		会話条件(会話の通し番号＋会話フォルダ番号＋会話の特徴を表す名前)：053-04　同性同士の断りの電話会話(対先輩)(女女)		話者記号の凡例： JFB0011：Japanese Female Base　001 JFO001：Japanese Female Older　001
	NCRB番号：-	会話時間：00:01:54	話者の数：2		
3	ライン番号	発話文番号	発話文終	話者	発　話　内　容
7	4	3	*	JFO001	うん。
8	5	1-3	/	JFB001	朝の9時にね,,
9	6	4	*	JFO001	うん=。
10	7	1-4	/	JFB001	=国立国語研究所に行ってね,,
11	8	5	*	JFO001	うん。
12	9	1-5	/	JFB001	私の代わりに,,
13	10	6	*	JFO001	うん。
14	11	1-6	*	JFB001	言語調査に関する実験に参加してはいただけないでしょうか?<軽く笑いながら>。
15	12	7	*	JFO001	えー、へー、へ、えー[あまりに驚いた様子]=。
16	13	8	*	JFB001	=しかも、韓国人と一緒に<笑いながら>。

Data：BTSJ Corpus

An example of dialogue (translated)

Speaker	Utterance
1	I called because I have a request.
2	Uh-huh.
1	Next Monday.
2	Uh-huh.
1	At 9 AM.
2	Uh-huh.
1	You need to go to NINJAL Lab.
2	Uh-huh.
1	On my behalf.
2	Uh-huh.
1	Would you be willing to participate in an experiment related to language research?
:	:

Estimation of Rx and Politeness Level by LLMs

- LLM estimated the value of Rx defined in Politeness Theory and the politeness level (Po) of utterances.

Speaker	Utterance
1	I called because I have a request.
2	Uh-huh.
1	Next Monday.
2	Uh-huh.
1	At 9 AM.
2	Uh-huh.
1	You need to go to NINJAL Lab.
2	Uh-huh.
1	On my behalf.
2	Uh-huh.
1	Would you be willing to participate in an experiment related to language research?
:	:

Rx	Po
2	3
3	3
1	3
1	4

$$W_x = D(S,H) + P(H,S) + R_x \quad (1)$$

W_x	Degree of face threat, indicating how much an act (x) threatens the listener's face.
D	Social distance between the speaker (S) and the hearer (H).
P	Power of the hearer (H) over the speaker (S).
R_x	Rank of imposition: cultural weight assigned to an act (x) based on the burden it places on the listener.

Each utterance was input into a LLM to estimate Rx and Po in five score in Likert scale.

Distance and Power

- Distance and Power are set by hand based on the relationship between speakers.
  - known / unknown
  - teacher / student
  - younger / older

D		P	
known	3	unknown	3
unknown	5	friend	3
		Speaker is older	4
		Speaker is younger	2
		Speaker is teacher	5
		Speaker is student	1

$$W_x = D(S,H) + P(H,S) + R_x \quad (1)$$

W_x	Degree of face threat, indicating how much an act (x) threatens the listener's face.
D	Social distance between the speaker (S) and the hearer (H).
P	Power of the hearer (H) over the speaker (S).
R_x	Rank of imposition: cultural weight assigned to an act (x) based on the burden it places on the listener.

Estimation of Rx and Politeness Level by LLMs

- Result of estimation

Speaker	D	P	Rx	Wx	Po	Utterance
1	3	4	2	9	3	I called because I have a request.
2	3	2	1	6	1	Uh-huh.
1	3	4	1	8	3	Next Monday.
2	3	2	1	6	1	Uh-huh.
1	3	4	1	8	3	At 9 AM.
2	3	2	1	6	1	Uh-huh.
1	3	4	3	10	3	You need to go to NINJAL Lab.
2	3	2	1	6	1	Uh-huh.
1	3	4	4	11	3	On my behalf.
2	3	2	1	6	1	Uh-huh.
1	3	4	4	11	5	Would you be willing to participate in an experiment related to language research?
2	3	2	1	6	1	Eh, huh, heh, eh.

Estimation of Rx and Politeness Level by LLMs

- Result of estimation

Speaker	D	P	Rx	Wx	Po	Utterance
1	3	4	2	9	3	I called because I have a request.
2	3	2	1	6	1	Uh-huh.
1	3	4	1	8	3	Next Monday.
2	3	2	1	6	1	Uh-huh.
1	3	4	1	8	3	At 9 AM.
2	3	2	1	6	1	Uh-huh.
1	3	4	3	10	3	You need to go to NINJAL Lab.
2	3	2	1	6	1	Uh-huh.
1	3	4	4	11	3	On my behalf.
2	3	2	1	6	1	Uh-huh.
1	3	4	4	11	5	Would you be willing to participate in an experiment related to language research?
2	3	2	1	6	1	Eh, huh, heh, eh.

Rx increases or decreases according to the burden an utterance places on the hearer.
Wx also increases or decreases in response to changes in Rx.

Estimation of Rx and Politeness Level by LLMs

- Result of estimation

Speaker	D	P	Rx	Wx	Po	Utterance
1	3	4	2	9	3	I called because I have a req
2	3	2	1	6	1	Uh-huh.
1	3	4	1	8	3	Next Monday.
2	3	2	1	6	1	Uh-huh.
1	3	4	1	8	3	At 9 AM.
2	3	2	1	6	1	Uh-huh.
1	3	4	3	10	3	You need to go to NINJAL Lab
2	3	2	1	6	1	Uh-huh.
1	3	4	4	11	3	On my behalf.
2	3	2	1	6	1	Uh-huh.
1	3	4	4	11	5	Would you be willing to parti research?
2	3	2	1	6	1	Eh, huh, heh, eh.

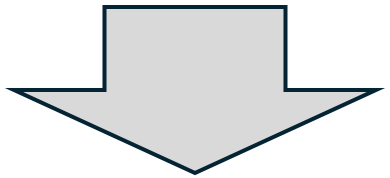
Politeness levels are appropriately assigned according to the content of the utterance.

The estimated values of Rx and politeness are subjective, and it is not easy to determine the accuracy of these estimates. After confirming that the results are reasonably intuitive, we proceeded to verification of politeness theory using these estimates.

Verification of Politeness Theory

- For all 25,844 utterances contained in 120 dialogues, we estimated the values of Rx and the politeness level (Po).
- The value of Wx was calculated as the sum of D, P, and Rx, in accordance with its definition.

Politeness is viewed as a linguistic strategy to compensate for face-threatening acts.



Wx values and politeness levels are correlated.

Verification of Politeness Theory

Correlation between Wx and Po
("#Dialog*" represents the number of dialogues where Wx and Po are significantly correlated)

Dialogue Category	#Dialog	#Dialog*	Avg. Wx	Avg. Po	Correl Coeff
Casual, Friends (Male-Male)	10	9	7.63	2.31	0.56
Casual, Friends (Female-Female)	21	19	7.66	2.29	0.70
Casual First-time Meeting (Female-Female)	11	11	9.59	3.38	0.42
Thesis Guidance (Teacher-Student)	10	8	7.98	3.32	0.18
Refusal (To Senior; Female-Female)	13	12	7.06	2.84	0.69
Refusal (To Peer; Female-Female)	13	10	7.59	2.73	0.57
Refusal (To Junior; Female-Female)	13	4	8.10	3.19	0.10
Request Between Friends (Male-Male)	10	7	7.49	2.66	0.58
Request Between Friends (Female-Female)	10	9	7.44	2.61	0.53
Debate, Between Friends (Mixed-Gender)	5	5	7.78	2.66	0.75
Debate, First-time Meeting (Female-Female)	4	4	10.33	3.49	0.12
	120	98 (81.7%)			

In 81.7% of the dialogues, Wx and politeness showed a significant correlation.

Verification of Politeness Theory

Correlation between Wx and Po
("#Dialog*" represents the number of dialogues where Wx and Po are significantly correlated)

Dialogue Category	#Dialog	#Dialog*	Avg. Wx	Avg. Po	Correl
Casual, Friends (Male-Male)	10	9	7.63	2.31	0.56
Casual, Friends (Female-Female)	21	19	7.66	2.29	0.70
Casual First-time Meeting (Female-Female)	11	11	9.59	3.38	0.42
Thesis Guidance (Teacher-Student)	10	8	7.98	3.32	0.18
Refusal (To Senior; Female-Female)	13	12	7.06	2.84	0.69
Refusal (To Peer; Female-Female)	13	10	7.59	2.73	0.57
Refusal (To Junior; Female-Female)	13	4	8.10	3.19	0.10
Request Between Friends (Male-Male)	10	7	7.49	2.66	0.58
Request Between Friends (Female-Female)	10	9	7.44	2.61	0.53
Debate, Between Friends (Mixed-Gender)	5	5	7.78	2.66	0.75
Debate, First-time Meeting (Female-Female)	4	4	10.33	3.49	0.12

Comparison of estimated politeness on relationship of speakers

Verification of Politeness Theory

Correlation between Wx and Po
("#Dialog*" represents the number of dialogues where Wx and Po are significantly correlated)

Dialogue Category	#Dialog	#Dialog*	Avg. Wx	Avg. Po	Correl
Casual, Friends (Male-Male)	10	9	7.63	2.31	0.56
Casual, Friends (Female-Female)	21	19	7.66	2.29	0.70
Casual First-time Meeting (Female-Female)	11	11	9.59	3.38	0.42
Thesis Guidance (Teacher-Student)	10	8	7.98	3.32	0.18
Refusal (To Senior; Female-Female)	13	12	7.06	2.84	0.69
Refusal (To Peer; Female-Female)	13	10	7.59	2.73	0.57
Refusal (To Junior; Female-Female)	13	4	8.10	3.19	0.10
Request Between Friends (Male-Male)	10	7	7.49	2.66	0.58
Request Between Friends (Female-Female)	10	9	7.44	2.61	0.53
Debate, Between Friends (Mixed-Gender)	5	5	7.78	2.66	0.75
Debate, First-time Meeting (Female-Female)	4	4	10.33	3.49	0.12

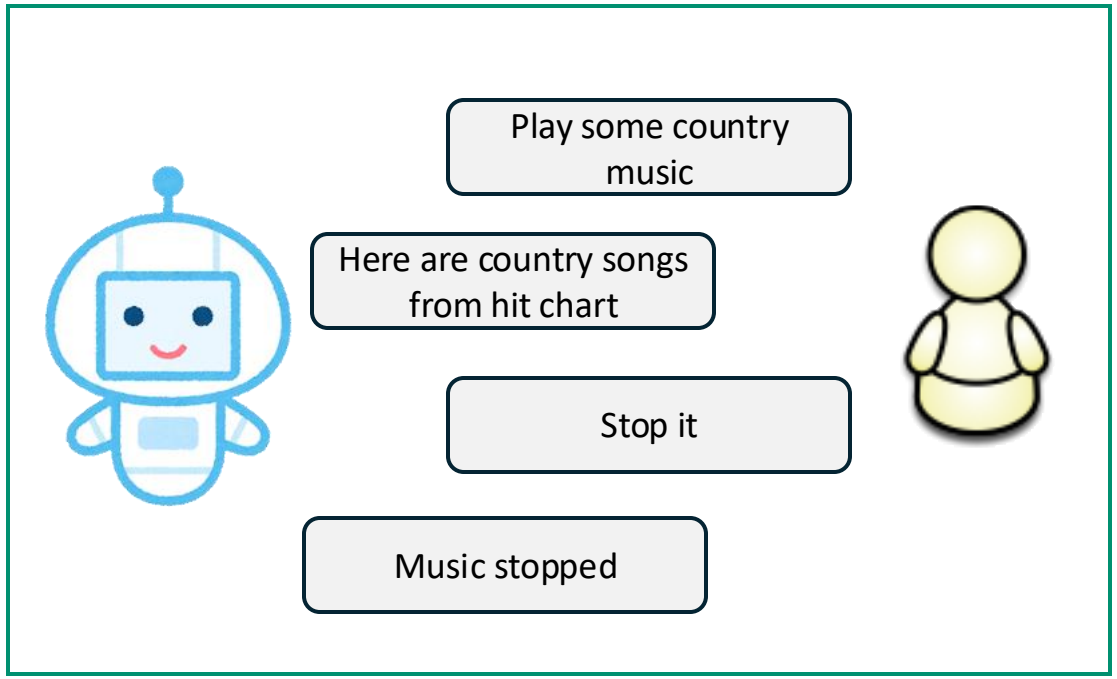
Comparison of estimated politeness on the topic of conversation

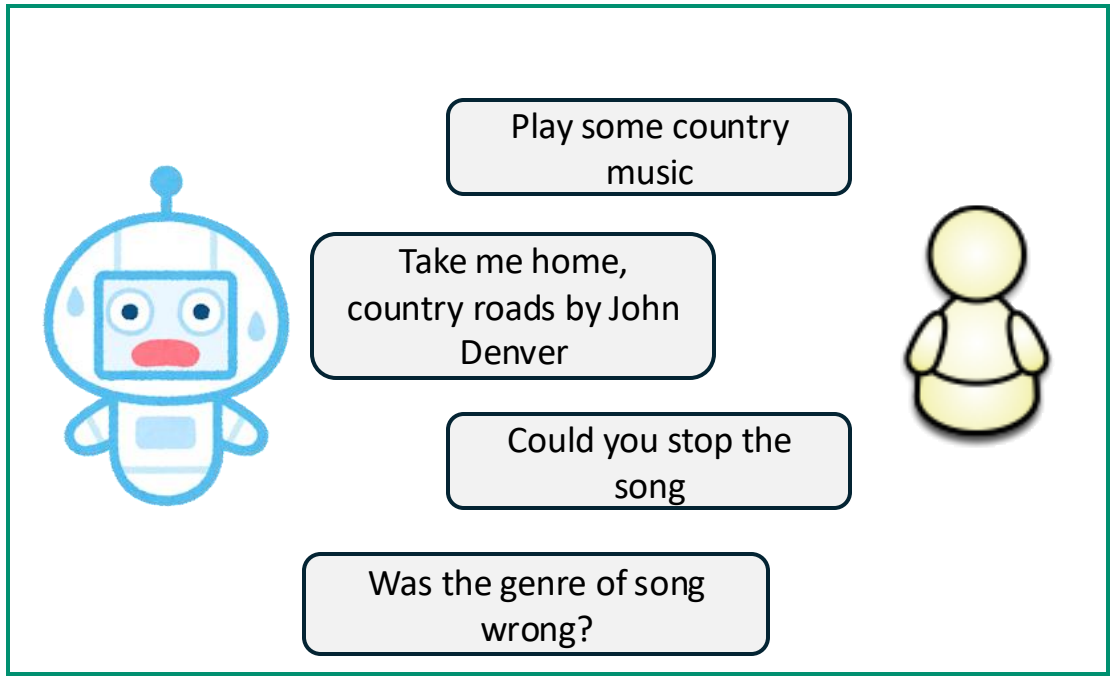
Conclusion

- Verification Results
  - Using the quantitatively estimated values of Rx (the degree of burden an utterance places on the hearer) and politeness level obtained through LLMs, we investigated the correlation between Brown and Levinson's formula for estimating face-threatening acts and politeness. We confirmed that **in 98 out of 120 dialogues**, the two were **significantly correlated**.
- Future Directions: Application in Dialogue Systems
  - Estimate Rx and politeness levels of user utterances, examine their relationship, and consider the possibility of implicit intentions if a discrepancy exists.
  - Instead of generating system utterances with a fixed level of politeness, deliberately adjust politeness according to social distance, power relations, and the degree of burden imposed by the utterance.

Application Ideas for Dialogue Systems

Estimate the Rx and politeness level of user utterances, examine their relationship, and consider the possibility of implicit intentions if a discrepancy exists between them.





Application Ideas for Dialogue Systems

- When generating system utterances, do not maintain a fixed level of politeness. Instead, deliberately adjust the politeness level according to social distance, power relations, and the degree of burden imposed by the utterance.