Modeling hyperpoliteness in the Russian language

Rhea Kapur (rheak@stanford.edu)
Stanford University

Abstract

Politeness is an essential part of how humans use language to convey respect for others. However, when speech is *overly* polite, the effect may instead be offensive. Linguistic theories, universals, and computational models concerning politeness (and to a lesser extent, hyperpoliteness) do exist, but they for the most part only focus on how the phenomena presents in English. To my knowledge, cross-cultural pragmatic differences in politeness strategies have only been theoretically formalized for the Russian language (Mills, 1992); even still, these differences have not yet been expressed in a formal computational model that accounts for how Russian-speaking peoples rely on background and context when producing different levels of polite speech. Here I present such a computational framework, modeling hyperpoliteness in Russian as an instance of ironic speech with various linguistic markers and with four potential affects: state, valence, arousal, and authority. I show that the model's characterizations of rude, polite, and hyperpolite speech closely match the reasonable and expected human interpretation and discuss broader implications on cross-cultural perceptions of politeness.

Keywords: politeness; hyperpoliteness; Russian; pragmatics; computational modeling; Slavic linguistics

Introduction

In conversation, humans choose to be polite for a variety of reasons: to convey appreciation, to be perceived as kind, to communicate respect for authority and any present power dynamics, to minimize threats, to fit in with social norms, and more. As acknowledged in Yoon et. al. (2018), polite speech often involves exaggeration, conditionals, overly elongated and dawdling utterances, and other unnecessary modifiers. As such, the nature of polite speech directly conflicts with Grice's four maxims of conversation, which assume that speakers abide by a succinct and straightforward manner of communication (Grice, 1975). Enter hyperpoliteness, or utilizing *overly* polite speech to either doubly appease or subtly offend a listener, which only further strays from the Gricean maxims but still plays an important role in conversation, especially between people in different relative positions of power. This presents an interesting challenge for the language understanding and information modeling task around politeness.

The spectrum of polite speech (rude, polite, hyperpolite) has been most extensively explored in context of the English language; there, theory and universals (Brown & Levinson, 1987; Baider et. al., 2020) as well as computational models (Yoon et. al. 2018) of the phenomena exist. However, cross-cultural pragmatic differences in polite speech remain for the most part undefined and unexplored. A notable exception is "Conventionalized Politeness in Russian Requests: A Pragmatic View of Indirectness," published by Margaret Mills in 1992 in *Russian Linguistics*. To my knowledge, this is the sole existing paper examining differences in the presentation of polite speech between English and conversational Russian. Mills details how politeness manifests oppositely in English and Russian; as per one of her examples, the expression "Couldn't you (possibly) pass the salt, could you?" would be considered rude in English, but its Russian equivalent—Вы не разменяне нямь рублей?—is incredibly polite and respectful. Specifically, Mills outlines various "politeness strategies" in Russian whose presence indicates polite speech: the markers ножалуцсма and просмуме as well as the future auxiliary будем are some, as well as use of the conditional tense and negation. Mills implies that with hyperpolite requests and utterances in Russian, an excess of politeness strategies are present (for instance, negation and the future auxiliary would be considered hyperpolite in contrast to just negation, which is solely considered polite). This general pattern is also confirmed by Baider et. al. (2020), which describes hyperpoliteness as "characterized by the massive presence of politeness markers." This proves especially important when considering the task of computationally modeling hyperpoliteness in the Russian language—the one aspect that Mills did not cover in her work and the main goal of this paper.

To generate a computational framework for hyperpoliteness, its relationship with irony must be considered. In fact, hyperpoliteness is better understood and modeled as an ironic rather than literal instance of politeness. The plain meaning of a hyperpolite utterance can in some cases be different than the speaker's intention; being overly polite can actually mean one is not being polite at all if, for instance, they wanted to express dissatisfaction with the listener but, as a subordinate, would harm their reputation by doing so with a direct and obviously pointed statement. Hyperpoliteness in Russian is best represented as a variation of the Kao. et. al. (2015) model of ironic speech that also accounts for the cross-cultural differences described in Mills (1992) around the definition of politeness in Russian as well as an extended set of goals for the speaker around what their intentions with the hyperpolite utterance may be. This is the approach I take with my model, which is described in further detail in the next section. I then

go on to show that the model's interpretations of rude, polite, and hyperpolite speech closely match the expected human interpretations, and I close with next steps and broader considerations in cross-cultural models of politeness and more.

Computational Model

My model of hyperpoliteness in the Russian language extends Kao et. al. (2015)'s irony model; understanding the latter first is critical, and so I present an overview here. The Kao irony model builds off of the general RSA (Rational Speech Act) framework, which represents language understanding and pragmatic reasoning as a recursive, back-and-forth exchange between the speaker and the listener with the overall goal of interpreting an utterance (Frank, 2016; Frank & Goodman, 2012). More specifically, and following the model of hyperbole presented in Kao, Wu et. al. (2014), the Kao irony model recognizes how the speaker may have conversational goals (termed QUDs, or questions of discussion) that are unknown to the listener; in addition to processing the meaning of an utterance, the listener must interpret those goals (see Kao et. al., 2015 for a more detailed explanation and for examples).

Structurally, the Kao irony model involves a literal listener $L_{literal}$, which reasons about a state s and the speaker's QUD surrounding s through literal interpretation, a speaker S which generates an utterance based on a distribution that takes into account the results of $L_{literal}$ in context of all possible QUDs, and a pragmatic listener $L_{pragmatic}$ considers the utterance returned by S and marginalizes over all possible QUDs. Kao et. al. define their literal interpretation function around whether the utterance is equal to the state, and they consider 3 QUDs (lend to priors over which conditioning occurs in the model): the speaker desires to communicate around s, the state of the world, arousal, or the strength of feeling about the state, or valence, i.e. whether the speaker feels positively or negatively about the state. Kao et. al. state that the resulting probability distribution over s and the potential QUDs reflects a final interpretation of the utterance.

Formalizing hyperpoliteness in Russian lends beautifully to not only the RSA framework—Mills (1992) already considers Russian requests under a very similar speaker-hearer (i.e., listener) speech act framework focused on a sole statement—but also to a modified and extended version of the Kao irony model with redefined priors/QUDs and a more specific and encompassing literal interpretation function; as mentioned previously, hyperpoliteness is considered an ironic use of politeness (meaning does not always match up with intention) and therefore the principles and structure of the Kao model apply.

In my model, there are four given utterance types, each represented by a placeholder and reflecting different levels of polite speech. The first, `Placeholder1`, corresponds to rude speech and includes none of the politeness strategies delineated in Mills (1992) and summarized in the introduction. The second, `Placeholder2`, corresponds to polite speech and includes an instance of the conditional tense as its only politeness strategy. The third, `Placeholder3`, also corresponds to polite speech and includes an instance of negation as its only politeness strategy. And finally, the fourth, `Placeholder4`, corresponds to hyperpolite speech and includes both an instance of the conditional tense *and* an instance of negation as the politeness strategies present. See Figure 1 for sample polite utterances in Russian from Mills (1992). `Placeholder2` corresponds to Sample 1, and `Placeholder3` corresponds to Sample 2. See Figure 2 for sample hyperpolite utterances in Russian (`Placeholder4`) from Mills (1992). See Figure 3 for sample rude utterance in Russian (`Placeholder1`) from Mills (1992).

1) *Я бы хотел, чтобы вы вернули мою книгу.*
I would like you to return my book, please.
[S's wish/want for H to perform A]

2) *Вы не можете вернуть мою книгу?*
Could you/can you please return my book?
[S's concern about H's ability to perform A]

Figure 1: Samples of polite (+ conditional/negation) utterances in Russian from Mills (1992).

a. *Не могли бы вы разменять мне пять рублей?*
b. *Не можете ли вы разменять мне пять рублей?*

Figure 2: Samples of hyperpolite (both conditional and negation present) utterances in Russian from Mills (1992). The first translates to "Could you change five rubles for me?" and the second translates to "Can't you change five rubles for me?"

*e. *Вы можете разменять мне пять рублей?*

Figure 3: Sample of rude or simply impolite (no syntactic marker or politeness strategy present) utterance in Russian from Mills (1992). This request roughly translates to "Can you change five rubles for me?" but the general construction (separate from Mill's politeness strategies) seems to indicate that the simple task of changing the five rubles will be difficult for the listener.

In terms of the model's priors (and subsequently QUDs), I redefine the state s to reflect how the speaker feels about the listener as a person (relating to their integrity, intelligence, personality, etc.), with 3 possible ordered values: `poor`, `neutral`, and `good`; this reflects how the human choice to be rude, polite, or hyperpolite directly relies on this state.

Valence remains whether the speaker feels positively or negatively about the state (so, whether the speaker feels positively or negatively about how they feel about the listener), and arousal too still reflects the speaker's strength of feeling about the state (so, the speaker's strength of feeling about how they feel about the listener) as in the standard irony model. Finally, I added one last prior around authority to account for how the relationship between the speaker and the listener affects the speaker's decision to produce a rude, polite, or hyperpolite utterance. The possible ordered values or relationships for this prior are formalized in Mills (1992) under a Speaker-Hearer Gradient, which "refers to the relative degree of power and/or authority perceived to exist between the two speech participants." They are: `S>L` (speaker is the superior, listener is the subordinate), `S=L` (speaker and listener are equals), and `S<L` (speaker is the subordinate, listener is the superior). The potential QUDs in the politeness model are: communicating about the state, communicating about the valence, communicating about the arousal, or communicating about the authority.

Considering the ordered values of both the authority and state priors in context of the selected utterance type gives the four cases accounted for in my modified literal interpretation function. The combination of a rude utterance (`Placeholder1`), a `poor` state, and a `S>L` authority makes sense and should be interpreted literally; a superior can afford to be rude to a subordinate if they feel poorly about them and not fear for their reputation. Additionally, the combination of a polite utterance (`Placeholder2` or `Placeholder3`) and a `neutral` state should also be interpreted literally; logically, standard politeness is the average human's safe bet in conversation if no strong feelings about the listener are present. Finally, a hyperpolite utterance (`Placeholder4`) with either a `poor` *or* `good` state and `S<L` authority makes sense; either the speaker, who is a subordinate position, feels positively about the listener and wants to especially and noticeably express this by being hyperpolite, *or* the speaker feels negatively about the listener, but cannot directly express this for fear of harming their reputation and perception from an important person and so resorts to being ironic and subtly offensive through hyperpoliteness. Literal interpretation is only defined to evaluate to true in these specific scenarios.

Evaluation

Running the politeness model will give the probability that a specific type of utterance makes sense given what was sampled from the priors. In this section, I will run three experiments (examining the four utterances) and analyze the resulting probabilities.

Rude Utterance

The tabular visualization of passing a rude utterance (`Placeholder1`) to $L_{pragmatic}$ is presented in Table 1.

Table 1: Probability distribution for rude utterance type.

state	valence	arousal	authority	probability
poor	-1	high	S > L	22%
poor	-1	high	S < L	16%
poor	-1	high	S = L	16%
good	1	high	S > L	11%
neutral	-1	low	S > L	6%
good	1	high	S < L	5%
good	1	high	S = L	5%
neutral	1	low	S > L	4%
neutral	-1	low	S < L	2%
neutral	-1	low	S = L	2%

The highest probability here as to whether the rude utterance makes sense given the selected priors is 22%, and this occurs when the speaker feels poorly toward the listener, their feeling about the state is negative, their strength of feeling about how poorly they feel toward the listener is high, and they are in a superior position to the listener. This makes sense; the speaker can afford to be rude in this case, especially because they are in a position of authority. The fact that the ordered value of `poor` for the state produces the highest probability is telling and evidences that the model matches expected behavior, because a rude utterance does make the most sense in that scenario.

Polite Utterance

The tabular visualization of passing a polite utterance (either `Placeholder2` or `Placeholder3`) to $L_{pragmatic}$ is presented in Table 2.

Table 2: Probability distribution for polite utterance type.

state	valence	arousal	authority	probability
neutral	1	low	S = L	10%
good	1	high	S = L	9%
neutral	-1	low	S = L	9%
poor	-1	high	S = L	8%
neutral	1	low	S < L	8%
neutral	1	low	S > L	8%
neutral	-1	low	S < L	7%
good	-1	low	S > L	7%
good	1	high	S < L	6%

Notice here how the probabilities are fairly even split across all the different combinations of values for priors. This makes sense; a standard, pleasant level of politeness is a safe bet in human conversation and can work in all scenarios. Additionally, notice how the combinations where the authority prior has value `S=L` do still rise to the top, with the highest priority; this reveals how the standard level of politeness still does make the most sense among equals. Additionally, notice how the state prior rarely takes on a `poor` value with these

top priorities; being polite does not make as much sense if the speaker feels poorly about the listener. Instead, and as I will show in the next section, hyperpoliteness is preferred.

Hyperpolite Utterance

The tabular visualization of passing a hyperpolite utterance (`Placeholder4`) to $L_{pragmatic}$ is presented in Table 3.

Table 3: Probability distribution for hyperpolite utterance type.

state	valence	arousal	authority	probability
good	1	high	S < L	19%
good	1	high	S = L	15%
good	1	high	S > L	15%
poor	-1	high	S < L	13%
poor	-1	high	S = L	8%
poor	-1	high	S > L	8%
neutral	1	low	S < L	4%
neutral	-1	low	S < L	4%
good	1	low	S < L	2%

Here, the highest probability as to whether the hyperpolite utterance makes sense given the selected priors is 19%, and this occurs when the speaker feels good about the listener, feels very strongly about that feeling, *and* is subordinate to that listener. This makes sense; the speaker would be hyperpolite as to really convey their affection and appreciation for the listener. Also of note is that another high probability (13%) occurs when the speaker feels poorly about the listener, but is still in a subordinate position, and so chooses to be hyperpolite as to subtly convey their dissatisfaction while still saving face. These two sensible scenarios are represented in the table and evidence that the model is working as expected. Notice as well how the highest probabilities with hyperpoliteness occur when the values for the different priors are on the farthest ends of their respective spectrums: `good` and `poor`, `high` and `low`, etc. This reveals how hyperpoliteness makes the most sense when the speaker feels very strongly in a particular direction about the listener. Unlike regular politeness, it is not a safe bet; in fact, it is a polarizing type of speech!

Additionally, the model favors the "overly nice" combination over the "subtly offensive" one, as evidenced by how the former has higher probabilities than the latter even though both are deemed equal by literal interpretation. This confirms how due to the risky nature of hyperpoliteness, the best bet is to err on the side of caution.

Discussion

In this paper, I formalized the theory surrounding politeness and hyperpoliteness in Russian into an RSA model that builds upon and extends the RSA model of irony presented in Kao et. al. (2015). I demonstrated the need for an additional authority prior in modeling and considering hyperpoliteness as an ironic form of politeness, a differing approach to Yoon et. al. (2018), and I analyzed how the probability distributions for each utterance type match the expected results based on human reasoning. Kao. et. al (2015) acknowledges how their minimal extension highlights the relationship between hyperbole and irony, and I too make a parallel conclusion here: the extension from irony to (hyper)politeness is equally minimal and only further serves to highlight the relationship between the two types of speech.

The model I presented in this paper describes the probability that a particular utterance matches samples from priors; it does not, however, model the probability that a speaker would choose to put forth a rude, polite, or hyperpolite utterance given samples from the priors. The latter is also an interesting scenario to consider. I envision the construction of such a model to be a hybrid of the politeness model described in Yoon et. al. (2018) and the irony model presented in Kao et. al. (2015). There would be two listeners and two speakers present, following Yoon et. al. (2018). $L_{literal}$ would be the first listener and implemented exactly the same as described here except cached. The first speaker *S* will recursively reason and infer about $L_{literal}$ with respect to both the authority *and* an additional prior for the listener's expected feelings. Implicitly taking into account conditionals, negatives, and other politeness strategies and syntactic markers, this prior represents how the speaker's utterance would make the listener feel if taken literally (`-1` for rude, `0` for polite, and `1` for hyperpolite, where the more negative values reflect the listener feeling worse). The authority prior reflects the informational context detailed in Yoon et. al. (2018), while the listener feeling prior reflects the social context; together with the constant `phi`, which reflects the extent to which the speaker is communicating to make the listener feel a certain way versus to convey a literal meaning, the two form a utility function that *S* takes into account when conditioning.

Next (still following Yoon et. al., 2018), $L_{pragmatic}$, or the second listener, would become a "joint-inference model" reasoning about *S* given an entire distribution over `phi` that reflects all possible levels of uncertainty surrounding `phi` (and subsequently *S*'s intentions with the utterance). Finally, the second speaker and final inference model *S2* would take effect, redefining the information and social components of *S*'s utility function to be based on $L_{pragmatic}$ instead of $L_{literal}$ and also adding in a final self-presentational component reflecting $L_{pragmatic}$'s perception of `phi` and revealing *S2*'s relative indirectness (Yoon et. al., 2018).

The result of adding in this additional layer completely transforms how the probability should be interpreted, as mentioned before; now, we model what politeness level the average speaker would choose in a given scenario outlined by sampling priors from the four priors (state, valence, arousal, and authority, all of which are defined the same), which would be incredibly interesting to explore. However, I experimented with implementing this model and found that the Infer() function in $L_{pragmatic}$ fails to converge; whether this is simply an error in my implementation or a greater prob-

lem with the MCMC (Markov chain Monte Carlo) sampling technique requires further investigation. This is something I would love to explore and confirm in further research.

A closing note on modeling a cross-cultural occurrence of a well-known speech phenomena: this work would not have been possible without the starting theory found in Mills (1992) around the differences in linguistic presentation of politeness and hyperpoliteness in Russian and English. For languages for which such politeness theory is unpublished, the corresponding models that take into account any cross-cultural differences likely do not exist. As such, theoretical and informational research around the structure and presentation of different linguistic phenomena in different languages should first and foremost be encouraged.

Acknowledgments

I would like to thank Alvin Tan and Veronica Boyce for their troubleshooting help and advice on structuring the politeness model.

References

Brown, J., & Levinson, S. (Eds.). (1987). *Some universals in language use (rev. ed.)*. Cambridge University Press.

E. Yoon, N. G., M. Tessler, & Frank, M. (2018). Polite speech emerges from competing social goals.

F. Baider, G. C., & Claudel, C. (2020). Researching politeness: From the 'classical' approach to discourse analysis... and back. *Corpus Pragmatics*, *4*, 259–272.

Frank, M. (2016). Rational speech act models of pragmatic reasoning in reference games.

Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998–998.

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, 3, speech acts* (pp. 41–58). New York: Academic Press.

J. Kao, L. B., J. Wu, & Goodman, N. (2014). *Nonliteral understanding of number words.*

Kao, J., & Goodman, N. (2015). Let's talk (ironically) about the weather: Modeling verbal irony. In *Proceedings of the 37th annual meeting of the cognitive science society.* Austin, TX: Cognitive Science Society.

Mills, M. (1992). Conventionalized politeness in russian requests: a pragmatic view of indirectness. *Russian Linguistics*, *16*, 65–78.