Expressiveness from a Bayesian Perspective*

Akitaka Yamada
Georgetown University

Keywords : expressiveness, addressee-honorifics, Bayesian statistics, dynamic pragmatics, variation theory

1. Introduction

Expressive elements are linguistic units that performatively express the speaker's emotive attitude which are independent from the regular descriptive content of the sentence (*aka.*, the at-issue meaning), such as (anti)honorifics, slurs, and epithets (Potts 2003, 2007; Potts and Kawahara 2004; Sells and Kim 2007; Hom 2010; McCready 2014, to appear; Hedger 2013; Blakemore 2015; Cepollaro 2015; Croom 2015; Yoon 2015; Davis and McCready 2016; Yamada 2018; *amo.*). For example, the sentence in (1) has two independent messages. The at-issue meaning is given in (i); *i.e.*, Mary ate the apple. The expressive meaning in (ii) (in this case, the honorific meaning) is subsidiary information associated with this sentence but orthogonal to the speech act of reporting Mary's apple-eating event. Likewise, when we compare (2)a with (2)b, it is clear that they both share the same primary information about Akitaka's nationality (the at-issue meaning) but (2)a departs from (2)b in that it also encodes the speaker's negative attitude toward a group of people to which he belongs (the expressive meaning) (Cepollaro 2015).

(1) *Mary-ga ringo-o tabe-masi-ta.*
Mary-NOM apples-ACC eat-HON_A-PST
(i) 'Mary ate the apple.' (ii) 'I respect you.'

(2) a. Akitaka is a Jap. b. Akitaka is Japanese.

While the language system only provides categorical distinctions (e.g., 'Jap' vs. 'Japanese' or the plain form vs. the *des-mas* form in Japanese), the intensity of our feeling is continuous. The main goal of this paper is propose a pragmatic model that links (i) the discrete choice in our language and (ii) the underlying continuous, latent, extra-linguistic variable(s).

Previous studies have proposed that real number intervals are appropriate for the denotation of expressive meanings as well as the information stored in the context (Potts 2007; McCready 2014; *cf.*, Potts and Kawahara 2004). While such real-based approaches give us flexible accounts for the semantics and the pragmatics of expressive elements --- with McCready (2014) being the most recent, sophisticated model ---, there are some issues not extensively discussed in the extant approaches. The first issue is a conceptual one. Under the previous approaches, discourse participants are assumed to know a particular range as the semantics for these expressive elements. But it remains unclear how they know what the range should be. I will argue that a slight modification of some assumptions under the real-based approach can circumvent the criticism about acquisition of the denotation of an expressive element. The second issue is about the record of the expressiveness. Previous theories only compare the interval of the current sentence with the interval of the immediate context. But I will show that the information of expressiveness in much earlier contexts should not be dismissed. In improving these real-based approaches, I will propose that (i) we can adopt a categorical denotation for the semantics of expressive elements (ii) while maintaining the gradability in our feelings expressed in the context if we regard the context shift in expressiveness as a Bayesian update of a relevant parameter/parameters stored in the structured discourse context.

2. McCready's (2014) Proposal

The most elaborated version of the real-based approach is given by McCready (2014) who proposes that the appropriateness of expressiveness is determined by the compatibility of (i) an interval proposed by the sentence (a subinterval of [0,1]) and (ii) an interval given by the context (also a subinterval of [0,1]; the degree of expressiveness is reflected in the magnitude of the interval, *e.g.*, High $\sqsubseteq$ [.6, 1), Mid $\sqsubseteq$ [.3, .7], Low $\sqsubseteq$ [0, .4]). If there is an overlap, the sentence is judged appropriate in the given context. The context update is triggered when the utterance interval is not wider than the context interval. For example, suppose the given context has the interval of [.9, 1.0] and the sentence gives the interval [.0, .1]. Clearly, there is no overlap between the two, so the use of the sentence is inappropriate (*e.g.*, saying a casual word in a very formal setting) but the fact that the utterance interval is not wider than the context interval updates the context interval. For illustration's sake, she adopts the arithmetic average for the new context interval: $[(.9+.0)/2,\ (1.0+.1)/2] = [.4,\ .45]$. In contrast, if the sentence gives the interval [.5, .95], the context interval does not change because the utterance interval is wider (= less informative) than the interval of the immediate context.

3. Issues

Although the basic idea that expressive meanings dynamically update the structured discourse context is attractive, there is room for improvement.

The first issue comes from the assumption that each discrete linguistic expression is associated with a particular continuous, real-based range. Some researchers, while admitting the gradability or the continuity of the scale in the conceptual/extralinguistic system, hesitate to incorporate the gradability (*i.e.*, the real-based denotation) into the semantics (Portner et al. to appear). If we adopt [.75, .9] for the range of *des* for illustration purposes, this means that the meaning is different from [.75, .89] or [.750001, .9]. In this framework, a child born to the Japanese speaking society is expected to identify the interval for these expressions. Since, by definition, [.75, .9] is sharply distinguished from [.750001, .9], finding a unique range does not sound like a plausible task.

The second issue is the record of the expressiveness in the discourse. Under previous approaches, once the context interval is updated at the time of t_i, the information on what level of expressiveness *had been* maintained before t_i becomes inaccessible, which seems unlikely. To see this, compare the two scenarios below.

(Scenario A) Previously, the speaker A had produced sentences with low range of intervals, such as [.2, .5], [.3, .4], … and [.2, .3]. However, at a point, he shifts to a high register and the context interval of the immediate context is set to [.75, .8], for example.

(Scenario B) Previously, the speaker A had produced sentences with a relatively high range of intervals, such as [.9, 1.0], [.8, .9], … and [.7, 1.0]. And now the context interval is set to

[.75, .8].

In both scenarios, the context interval of the immediate context is the same. Suppose, however, that in the next move the same speaker A has produced a sentence whose utterance interval is [.3, .4]. If we only track the immediate context, what we take into account is the relation between the immediate context interval [.75, .8] and the utterance interval of the new sentence [.3, .4]. We would predict that the new sentence should be surprising in both cases. But if we are in Scenario A, the new utterance range is indeed a familiar register despite the fact that it is far from the context interval of the immediate context. If we correctly predict that the move in Scenario B is more surprising, we somehow want to look at the record of the expressive dimension in earlier contexts. Certainly, one can propose a discourse context such as in (3), where, in addition to our familiar components of the discourse context, we have $h^{(n)}$, which stores ALL the previous context intervals up to the current n-th utterance. But if our conversation continues, h becomes a lengthy list, suggesting we remember too many past states. Is there any better model in which the history of prior contexts sufficiently influences our decision in measuring the surprisingness or the appropriateness of the current utterance without serious memory overload?

(3) a. $c_n = \langle cg, qs, tdl, \ldots, h_n \rangle$

b. $h_n = \langle h^{(1)}, h^{(2)}, h^{(3)}, \ldots, h^{(n)} \rangle$

where $h^{(i)} \in \{[\alpha, \beta]: 0 \leq \alpha \leq \beta \leq 1\}$.

4. Proposal

We can kill two birds with one stone if we adopt some ideas from the Bayesian statistics. I maintain the assumption that the context is modeled as in (3)a, but instead of having a lengthy list of past expressive states in (3)b, in my model h is a set of a few representative parameters that summarize what the past situations were like. For simplicity's sake, I adopt a model with two summary parameters (for reasons stated below). Under this model, the context update is seen as a function from $h_{i-1}=(\alpha_{i-1},\beta_{i-1})$ to $h_i=(\alpha_i, \beta_i)$, leading to the architecture in (4).

(4) a. $c_n = \langle cg, qs, tdl, \ldots, h_n \rangle$

b. $h_n = (\alpha_n, \beta_n)$

However, the question remains of how to find a reasonable set of summary parameters that capture the past states. Below, I will propose that a Bayesian interpretation of the context update in expressiveness gives us a clue as to how to set up these parameters. To this end, we will begin with a brief introduction on the philosophy of this statistic framework and then discuss how the context parameters are updated.

4.1. Bayesian Statistics and Dynamic Semantics/Pragmatics

The main purpose of the inferential statistics is to make an inference about the parameters of a statistical model from the available data. In the Bayesian paradigm, (i) probability is used to measure any subjective uncertainty and, thus, (ii) parameters of a statistical model are given a probability distribution if we have some uncertainty about them; *cf.*, these assumptions are not made in non-Bayesian approaches.

For example, assume that we have a coin, which can be fair but may be biased. If we keep flipping the same coin, we will gain a sequence of outcomes and, based on them, we can infer with what probability the coin lands on heads.

Suppose the outcome of the coin toss follows

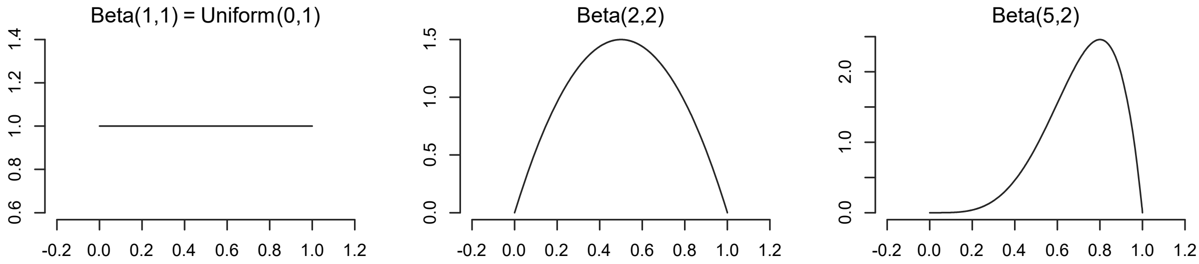

Figure 1. Modeling our uncertainty.

the Bernoulli distribution with an unknown parameter π. We will keep updating our estimation of this parameter every time we see the outcome. Since the uncertainty is captured in the form of a probability distribution in this framework, the estimation process is understood as an update of the probability distribution for π. For instance, before we see the results, we might have been completely agnostic about the possible value for π; it can be very small, *e.g.*, .05 or .08; it can be very high, *e.g.*, .97 or .94; or, it can be fifty-fifty. The probability distribution in the leftmost panel in Figure 1, for example, represents this agnostic state. But, after seeing the data, we update our uncertainty about π. If we observe one head and one tail, we may infer that this is a fair coin (the center panel in Figure 1, in which the values around .5 are most likely). But, if we have seen three heads in a row, we are more inclined to believe that π should take a high value, though it remains true that we never know the definite value for π (the rightmost panel in Figure 1).

We can understand this estimation process as a particular type of context update. Under the tradition of dynamic approaches to the discourse, the meaning of a sentence is regarded as a context change potential; the new context C' is derived from the previous context C and the current utterance S. This process is expressed as in (5)a. What the Bayesian statistics does is update our uncertainty with our new data --- which is expressed as a probability distribution as in Figure 1. If we refer to the previous uncertainty and the new uncertainty as U and U', and the data as D, the process of our estimation is expressed as in (5)b. The similarity is evident.

(5) a. C+S=C' b. U+D=U'

4.2. Context Shift in Expressiveness is a Bayesian Update

The main thesis of this paper is to adopt (5)b for the update mechanism in the expressive dimension of the discourse, which leads to the architecture in (4)b. The probability distribution is uniquely identified once we have specified the parameter(s) of that model. Thus, the nature of the update in (5)b is seen as the change of the values of these parameters. When we model a variable that ranges between 0 and 1, *i.e.*, in our case π, it is common to adopt the Beta distribution for its simplicity, which have two parameters α and β. Certainly, if we want to assume a more complicated model, a different set of summary parameters are used in place of (α, β). But, for simplicity's sake, I will keep using the Beta distribution and regard (α, β) as the summary parameters (= (4)b).

In the case of flipping coins, we estimate the parameter of the Bernoulli distribution π by observing the number of heads and tails. By analogy, we can interpret the pragmatics of addressee-honorification as an update of our subjective uncertainty parameter of the speaker's consistency of the *des-mas* form by observing how many *des-mas* forms we have heard in prior contexts. For example, when we meet a new person, we do not know how consistently this person uses an addressee-honorific marker when addressing us. So, again, the leftmost panel in Figure 1 is a fair starting distribution. There is uncertainty about this π. If this person keeps producing an addressee-honorific marker, we start getting a skewed distribution as in the rightmost panel in Figure 1. If this person wants us to have a good impression of him, he will try to produce as many utterances with *des*/*mas* as he can so that we end up having a very skewed distribution. In other words, the goal of producing an addressee-honorific marker is to let the hearer make a Bayesian update. By interpreting $h=(\alpha, \beta)$ as the set of parameters for the probability distribution for π (the parameters of the Beta distribution), the nature of the context update is seen as the change from the previous state $h_{i-1}=(\alpha_{i-1},\beta_{i-1})$ to the new state $h_i=(\alpha_i,\beta_i)$.

In our model, the semantic contribution of *des*/*mas* is essentially the same as the information that we gain when we see a head in the coin toss; *i.e.*, they are context change potentials updating α and β. In the coin-flipping setting, with little mathematics, it can be easily shown that (i) every time we see a head, we increment α by one and (ii) every time we observe a tail, we add one to β. The dynamicity of the context change triggered by the addressee-honorific marker and the plain form (which has no morphological realization; so I use ϕ to refer to this form) is defined as in (6).

(6) a. $h_{i-1} + [\![\textbf{\textit{des/mas}}]\!] = h_i$

where $h_{i-1}=(\alpha_{i-1},\beta_{i-1})$ and $h_i=(\alpha_{i-1}+1,\beta_{i-1})$

b. $h_{i-1} + [\![\,\phi\,]\!] = h_i$

where $h_{i-1}=(\alpha_{i-1},\beta_{i-1})$ and $h_i=(\alpha_{i-1},\beta_{i-1}+1)$

4.3. Solution to the Problems

Notice that we depart from the assumption of the interval-based approach that the denotation of expressive elements is an interval. In (6), the meaning of *des/mas* is an instruction to change the context parameter. No interval is assigned to the denotation of *des/mas*. This is a desirable consequence because our model is not sensitive to very small differences in real number intervals. As pointed above, in the previous interval-based approach, a new born child needs to identify the right intervals for *des* and *mas* and is expected to be sensitive to the difference between [.75, .9] and [.750001, .9]. On the other hand, our theory assumes the only thing a child is supposed to realize is that we add one to α when we encounters *des/mas*. Sensitivity to the difference between [.75, .9] and [.750001, .9] never arises in our semantics.

The other problem is sensitivity to the past conversation. The two parameters α and β succinctly and sufficiently summarize what the past expressive states were like. Since the presence of an addressee-honorific marker only increments the value of α, but not β, the magnitude of α reflects how much exposure we have to *des/mas* forms in the past. Likewise, the magnitude of β summarizes how often we have seen the plain form in earlier contexts. Though what we keep tracking are only two things, α and β, and not the entire conversation, we can still to some extent reconstruct what past interactions were like by interpreting the magnitude of these values.

Figure 2 displays some instances of Beta distributions with different parameter specifications. When α and/or β take a large value, the spread of the distribution gets smaller and smaller, suggesting we are confident in possible range for π.

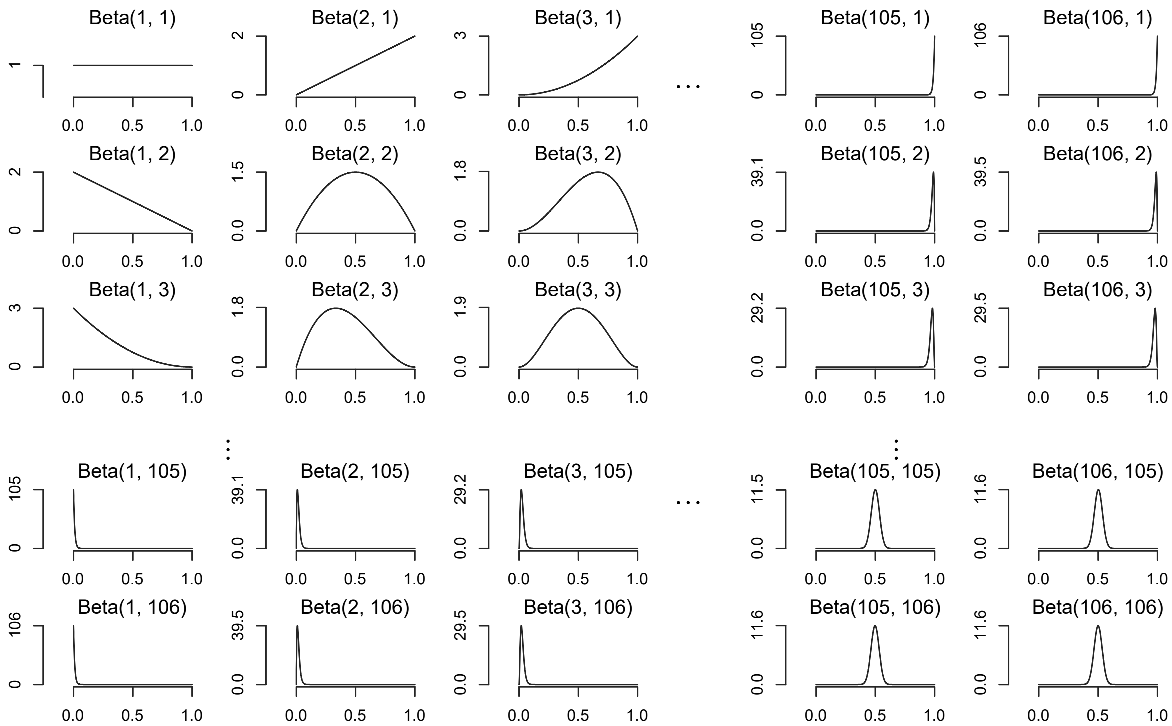

Figure 2. Beta distributions.

If we have enough exposure to, for example, plain forms, as assumed in Scenario A in Section 3, our expressive state is something like Beta (1, 105). Even though the speaker temporarily uses a few utterances with *des/mas*, that is, even if we update the expressive state to Beta (3, 105), we still have a relatively large value for β and thus we can predict, as expected, that the speaker will go back to his original tone and will consistently keep producing the plain form, as illustrated in the panel for Beta $(3, 105)$, whose distributional peak is still leaned toward to the left. Thus, even though we see a new utterance without *des/mas*, we will not be surprised at all.

In this way, we can quantitatively predict which form is more likely to be used in the next utterance without remembering all the past honorific/expressive states.

5. Conclusion and Future Directions

Inheriting the insight from previous studies that a specific component in the structured discourse context represents our real-based emotional state, this paper has proposed that the contextual update of such an expressive dimension (in our case *h*) is better understood as a Bayesian update. The denotation of expressive elements is seen as an instruction on how to change the parameters. No interval needs to be assumed for the denotation of expressive elements and the influence of non-immediate contexts is sufficiently captured.

In future studies, we can extend this analysis in many ways. First, we can elaborate the expressive state by taking into account different discourse participants. Alice's consistency in using *des/mas* to Bob would be different from Bob's consistency in *des/mas* to Alice. By replacing (4)b with the set of triples as in (7), we can track as many dimensions as we want.

(7) $$h_n = \left\{ \begin{array}{c} \langle Alice, (\alpha_n, \beta_n), Bob \rangle, \\ \langle Bob, (\gamma_n, \delta_n), Alice \rangle, \\ \vdots \end{array} \right\}$$

Second, we can also flesh out the structure for π. Alice's consistency in *des/mas* can be affected by many factors. Some are global indices such as the gender(s) of the participants, the social hierarchy and the age difference, while others are temporary, such as emotional state (*e.g.*, anger), power management, face-threat and sarcasm (McCready 2014; Brown 2015; Portner et al. to appear). Incorporation of such factors in the model is achieved by relating π_n with such variables via a link function, *e.g.*, $\mathrm{logit}(\pi_n)=\lambda_0+\lambda_{\mathrm{hier}}\, x_{\mathrm{hier},\, n} +\lambda_{\mathrm{age}}\, x_{\mathrm{age},\, n} +\ldots$.

This formula is equivalent to the Bayesian interpretation of the formula used in Variation Theory (Cedergren and Sankoff 1974). Attempts to improve the discourse model, thus, end up crossing a bridge between sociolinguistics and formal semantics/pragmatics, which have been conceived of as rather independent domains in linguistics (Yamada 2018).

REFERENCES

Blakemore, Diane (2015) "Slurs and Expletives: A Case against a General Account of Expressive Meaning," *Language Sciences* 52, 22-35.

Brown, Lucien (2015) "Honorifics and Politeness," *The Handbook of Korean Linguistics*, ed. by Lucien Brown and Jaehoon Yeon, 303-319, Wiley & Sons: US.

Cedergren, Henrietta J. and David Sankoff (1974) "Variable Rules: Performance as a Statistical Reflection of Competence," *Language* 50, 333-355.

Cepollaro, Bianca (2015) "In Defence of a Presuppositional Account of Slurs," *Language and Sciences* 52, 36-45.

Croom, Adam M. (2015) "Slurs, Stereotypes, and In-Equality: A Critical Review of 'How Epithets and Stereotypes are Racially Unequal'," *Language Sciences* 52, 139-154.

Davis, Christopher and E. McCready (2016) "Expressives in Questions," *SALT* 26, 1119-1138.

Hedger, Joseph A. (2013) "Meaning and Racial Slurs: Derogatory Epithets and the Semantics/Pragmatics Interface," *Language and Communication* 33, 205-213.

Hom, Christopher (2010) "Pejoratives," *Philosophy Compass* 5, 164-185.

McCready, E. (2014) "A Semantics for Honorifics with Reference to Thai," *PACLIC* 28, 503-512.

McCready, E. (to appear) *The Semantics and Pragmatics of Honorification*.

Portner, Paul, Miok Pak and Raffaella Zanuttini (to appear) "The Speaker-Addressee Relation at the Syntax-Semantics Interface," *Language*.

Potts, Christopher (2003) *The Logic of Conventional Implicatures*, PhD thesis, University of California, Santa Cruz.

Potts, Christopher (2007) "The Expressive Dimension," *Theoretical Linguistics* 33, 165-197.

Potts, Christopher and Shigeto Kawahara (2004) "Japanese Honorifics as Emotive Definite Descriptions," *SALT* 14, 235-254.

Kim, Jong-Bok and Peter Sells (2007) "Korean Honorification: A Kind of Expressive Meaning," *Journal of East Asian Linguistics* 16 (4), 303-336.

Yamada, Akitaka (2018) "Bayesian Update: from Semantics/Pragmatics to Variation Theory," presented at GURT.

Yoon, Suwon (2015) "Semantic Constraint and Pragmatic Nonconformity for Expressives: Compatibility Condition on Slurs, Epithets, Anti-Honorifics, Intensifiers, and Mitigators," *Language Sciences* 52, 46-69.