Communicative pressures at the semantics-pragmatics interface

(– draft August 9, 2016—)

1 The semantics-pragmatics divide

Abstract

Natural languages allow for the conventional pragmatic enrichment of semantic structure in a notably productive fashion. As a result, the information conveyed often goes beyond the literal meaning of an expression. This raises a challenge to explain the regular selection of particular semantics over others in light of pragmatics. To address this challenge, we propose a general model for the analysis of linguistic pressures that integrates (iterated) Bayesian learning in the replicator-mutator dynamics commonly used in evolutionary game theory. This model allows for population-level analyses of rational probabilistic language users with varied degrees of pragmatic sophistication and distinct languages. In an application to scalar implicatures, we show that semantics that allow for pragmatic enrichment are selected for when languages are pressured towards learnability. Crucially, this result hinges on pragmatics to compensate for the disadvantage in expressivity that users of such languages incur otherwise.

In linguistic theorizing, it is common to draw a distinction between semantics and pragmatics. Broadly speaking, the former concerns the truth-conditional content of expressions, whereas the latter concerns information beyond such literal meanings. For instance, the integration of context and the derivation of defeasible inferences. Thus, the information conveyed by an utterance is seldom, if ever, solely determined by semantics, but rather in tandem with pragmatics.

Much research in semantics and pragmatics is aimed at characterizing classes of expressions in terms of either domain, or their interplay. Thus, while there is disagreement on where their boundary lies, the distinction between semantics and pragmatics plays an important role in linguistics. However, an issue that has received little attention is the justification of semantic structure in light of pragmatics. That is, an account for the selection and pervasiveness of particular semantic structures over others under consideration of the informational enrichment provided by pragmatics.

A number of recent investigations have began to address issues pertaining to the development and selection of linguistic properties (see Steels 2015 and Tamariz and Kirby 2016 for recent overviews). Our starting point is given by the overarching account of competing pressures that has crystallized across these approaches: Natural languages are pressured for successful communication, one the one hand, and for acquisition, on the other. That is, they need to be well-adapted to the communicative needs within a linguistic community, as well as be learnable to survive their faithful transmission across generations.

We proceed by instantiating these pressures in the replicator-mutator dynamics, commonly used in evolutionary game theory. This model allows for general and precise means to model the dynamics of linguistic pressures by combining functional pressure on successful communication, effects of learning biases on (iterated) Bayesian learning (Griffiths and Kalish 2007), and probabilistic models of language use in populations with distinct lexica (Frank and Goodman 2012, Franke and Jäger 2014, Bergen et al. 2016). Drawing from the latter research strand allows us to make the distinction between semantics and pragmatics in communicative behavior precise. In this way, this model bridges the recent surge of synchronic probabilistic rational language use analyses with diachronic models of cultural evolution. We then analyze the prevalence of a lack of semantic upper-bounds in the literal meaning of weak scalar expressions. We show that a lack of upper-bounds is selected for when learners are biased towards simpler semantic representations, provided they have means to convey upper-bounds, i.e., provided that they can be derived via pragmatic reasoning.

2 Simplicity and expressivity at the semantics-pragmatics interface, and beyond

- General introduction to the idea of competing pressures in shaping languages (Zipf 1949, Piantadosi,...
- Idea in pragmatics: Clark and Wilkes-Gibbs 1986, Horn 1984
- Discuss Q- and R-predictions for language change (cf. Traugott 2004)
- Idea in LOT (Piantadosi et al. under review)

2.1 Competing pressures in cultural evolution

- discuss past research on interplay of pressures from a population-level perspective (see discussion in Kirby et al. 2015 for IL
- Learnability in iterated learning
- Expressivity in GT and some IL
- Sketch of replicator-mutator dynamics

More detailed discussion of models of cultural evolution. Short overview of past research with a focus on the difference between pure IL and functional pressure together with IL. **Possibly add a direct comparison of IL and RMD in the appendix using the setup of Griffiths & Kalish 2007. This may not be necessary.**

The emergence and change of linguistic structure is driven by many factors, from biological and socio-ecological to cultural (Steels 2011, Tamariz and Kirby 2016). Broadly put, social and ecological pressures determine communicative needs, while biology determines the architecture available for its use. Our focus is on the latter, cultural, factor, wherein linguistic structure is analyzed in terms of its use, as well as its transmission across generations.

As already mentioned in §1, research on selectional forces that apply in the cultural evolution of language has focused on two main pressures: expressivity and learnability. However, while it is generally acknowledged that both play a pivotal role, past approaches have focused exclusively, or at least emphasized, the role of one over the other (a recent exception is Kirby et al. 2015).

Expressiveness, or communicative efficiency, has been at the center of applications of evolutionary game theory to linguistics (Nowak and Krakauer 1999, Huttegger and Zollman 2013), explain RMD

In contrast, the iterated learning paradigm has focused on the effects of language transmission from generations of speakers to the next. explain IL

3 Model

3.1 Replicator-mutator dynamics

(II) Sequences and atomic observations. Before, the set of all observations was $O = \{\langle\langle s_1, m_i\rangle, \langle s_2, m_j\rangle\rangle | m_i, m_j \in M\}$. A member of O encodes that a teacher produced m_i in state s_1 and m_j in s_2, i.e., it encodes one witnessed message for each state. A datum d was a sequence of length k of members of O. Learners witnessed such data sequences. Now, more in line with Griffiths and Kalish (2007), $O = \{\langle s_i, m_j\rangle | s_i \in S, m_j \in M\}$ and d is a sequence of length k of members of O. The main difference is that now some d do not provide any production information for some states.

(III) Observations as production. Instead of taking the space of all possible sequences of length k into consideration, we take sample from O k-times according to the production probabilities of each type; $P(o = \langle s, m\rangle | t_i) = P(s)P(m|s, t_i)$. n such k-length sequences are sampled for each type. As a consequence, the data used for computing Q_i is not the same as that used for j $(i \neq j)$.

(IV) Parametrized learning $Q_{ij} \propto \sum_d P(d|t_i)F(t_j, d)$, where $F(t_j, d) \propto P(t_j|d)^l$ and $l = 1$ corresponds to probability matching and, as l increases towards infinity, to MAP.

The proportion of players of type i, x_i, is initialized as an arbitrary distribution over T. $p^\star \in \Delta(T)$ is learning a prior over (player) types dependent only on the lexicon of the type.

- $f_i = \sum_j x_j U(x_i, x_j)$
- $\Phi = \sum_i x_i f_i$
- $Q_{ij} \propto \sum_d P(d|t_i)\ P(t_j|d)$, where $P(t_j|d) \propto [P(t_j)P(d|t_j)]^l$, d is a sequence of observations of length k of the form $\langle\langle s_i, m_j\rangle, \ldots \langle s_k, m_l\rangle\rangle$, and $l \geq 1$ is a learning parameter.
- For parental learning (standard RMD): $\dot{x}_i = \sum_j Q_{ji} \frac{x_j f_j}{\Phi}$

3.2 Expressiveness as fitness-relative replication

Symmetrized expected utility. With $P \in \Delta(S)$ (uniform so far; $P = pr$):

- $U(t_i, t_j) = [U_S(t_i, t_j) + U_R(t_i, t_j)]/2$
- $U_S(t_i, t_j) = \sum_s P(s)\ \sum_m P_S(m|s; t_i)\ \sum_{s'} P_R(s'|m, t_j)\ \delta(s, s')$, where $\delta(s, s')$ returns 1 iff $s = s'$ and otherwise 0
- $U_R(t_i, t_j) = U_S(t_j, t_i)$

3.3 Iterated learning as acquisition-based mutation

3.4 Signaling behavior

Leave languages undefined until application

Signaling behavior. Exposition of signaling behavior as reasoning hierarchy, which we use to make a distinction between semantic and pragmatic language users

With $\lambda \geq 1$ (rationality parameter), $\alpha \in [0,1]$ (pragmatic violations) and $pr \in \Delta(S)$ a common prior over S (uniform so far):

$$R_0(s|m;L) \propto pr(s)L_{sm} \tag{1}$$
$$S_0(m|s;L) \propto \exp(\lambda\, L_{sm}) \tag{2}$$
$$R_1(s|m;L) \propto pr(s)S_0(m|s;L) \tag{3}$$
$$S_1(m|s;L) \propto \exp(\lambda\, R_0(s|m;L)^{\alpha}) \tag{4}$$

4 Lack of semantic upper-bounds in lexical meaning

Procedural description. The game is initialized with some arbitrary distribution over player types. At the game's onset we compute Q once based on the sets of sequences D (one for each parent type). Replicator dynamics are computed based on the fitness of each type in the current population as usual. Q is computed anew for each independent run (of g generations) given that it depends on D, which is sampled from production probabilities.

Languages. We consider a population of players with two signaling behaviors, literal and Gricean (level 0 and 1 below), each equipped with one of 6 lexicons. This yields a total of 12 distinct player types $t \in T$. $|M| = |S| = 2$, i.e., a lexicon is a $(2,2)$-matrix. These are listed in Table 1.

$$L_1 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \quad L_2 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad L_3 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$
$$L_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad L_5 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \quad L_6 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

Table 1: Set of considered lexica.

As in the CogSci paper, L_4 (semantic upper-bound for m_2) and L_5 (no semantic upper-bound for m_2) are the target lexica. Gricean L_5 users can convey/infer the bound pragmatically, while literal/Gricean L_4 users do so semantically.

4.1 Model parameters & procedure

1. Sequence length k
2. Pragmatic production parameter α
3. Rationality parameter λ
4. Learning prior over types (lexica); cost parameter c. $p^{\star}(t_i) \propto n - c \cdot r$ where n is the total number of states and r that of upper-bounded messages only true of s_1 in t_i's lexicon (if only s_1 is true of a message, then this message encodes an upper-bound). Then the score for L_1, L_3, L_5 is 2, that of L_4 and L_6 is $2-c$, and that of L_2 is $2-2c$; Normalization over lexica scores yields the prior over lexica (which is equal to the prior over types).

5. Prior over meanings (pr). We assume that $pr(s) = \frac{1}{|S|}$ for all s.
6. True state distribution (P). We currently assume that $P = \frac{1}{|S|}$ but it may be interesting to vary this
7. Learning parameter $l \geq 1$ with 1 corresponding to probability matching, and MAP as l approaches infinity
8. n is the sample of sequences of observations of length k sampled from the production probabilities of each type
9. Number of generations g

5 Discussion

6 Extensions

(I) Cost for pragmatic reasoning. At least in the CogSci setup the effect of adding cost to pragmatic reasoning is unsurprising: High cost for pragmatic signaling lowers the prevalence of pragmatic types. Lexica that semantically encode an upper-bound benefit the most from this. However, the cost needed to be substantial to make the pragmatic English-like lexicon stop being the incumbent type (particularly when learning is communal).

(II) Negative learning bias. Instead of penalizing complex semantics (semantic upper-bounds) one may consider penalizing simple semantics (no upper-bounds). This is useful as a sanity check but also yields unsurprising results in the CogSci setup: The more learners are biased against simple semantics, the more prevalent are lexica that semantically encode upper-bounds.

(III) Inductive bias. A second learning bias that codifies the idea that lexica should be uniform, i.e. be biased towards either lexicalizing an upper-bound for all weaker alternatives in a scalar pair or for none.

(IV) Uncertainty. The other advantage of non-upper bounded semantics lies in being non-committal to the negation of stronger alternatives when the speaker is uncertain. Adding this to the model requires the most changes to our present setup and some additional assumptions about the cues available to players to discern the speaker's knowledge about the state she is in.

(V) More scalar pairs. Taking into consideration more than one scalar pair. Preliminary results suggest that this does not influence the results in any meaningful way without further additions, e.g. by (III).

(VI) More lexica. Not necessary. Preliminary results suggest that considering more lexica has no noteworthy effect on the dynamics (tested with all possible 2x2 lexica).

(VII) State frequencies. Variations on state frequencies. This may have an interesting interaction with (III).

(VIII) Reintroduction of communal learning. One possibility: The probably N_{ij} with which a child of t_i adopts t_j could be the weighted sum of Q_{ij} (as before) and a vector we get from learning from all of the population: $L_j = \sum_d P(d|\vec{p})P(t_j|d)$, where $P(d|\vec{p}) = \sum_i P(d|t_i)\vec{p}_i$ is the probability of observing d when learning from a random member of the present population distribution.

7 Conclusion

References

Luc Steels. *The Talking Heads experiment: Origins of words and meanings.* Language Science Press, 2015.

Monica Tamariz and Simon Kirby. The cultural evolution of language. *Current Opinion in Psychology*, 8:37–43, 2016.

Thomas L. Griffiths and Michael L. Kalish. Language evolution by iterated learning with bayesian agents. *Cognitive Science*, 31(3):441–480, 2007.

M. C. Frank and N. D. Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998, 2012.

Michael Franke and Gerhard Jäger. Pragmatic back-and-forth reasoning. *Semantics, Pragmatics and the Case of Scalar Implicatures.*, pages 170–200, 2014.

Leon Bergen, Roger Levy, and Noah D Goodman. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 2016.

George Zipf. *Human behavior and the principle of least effort.* Addison-Wesley Press, 1949.

Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22 (1):1–39, 1986.

Laurence R. Horn. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin, editor, *Meaning, Form and Use in Context*, pages 11 – 42. Georgetown University Press, Washington, 1984.

Elizabeth Closs Traugott. Historical pragmatics. In Laurence R. Horn and Gregory Wand, editors, *The Handbook of Pragmatics*, pages 538–561. Blackwell Publishing, 2004.

Steven T. Piantadosi, Joshua B. Tenenbaum, and Noah D. Goodman. Modeling the acquisition of quantifier semantics: a case study in function word learnability, under review.

Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102, 2015.

Luc Steels. Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4):339–356, 2011.

M. A. Nowak and D. C. Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033, 1999.

S. M. Huttegger and K. J. S. Zollman. Methodology in biological game theory. *The British Journal for the Philosophy of Science*, 64(3):637–658, 2013.