Meaning evolution in multiple reference games with pragmatic language users

Michael Franke

1 Motivation

Standard game theoretic models of meaning evolution assume a single fixed *stage game*. What evolves are agents' strategies (ways of behaving) in that particular game. For example, we might consider a single signaling game with particular payoffs and ask what its equilibria are or which outcomes are frequently observed under a particular learning dynamic (Skyrms, 2010a; Huttegger and Zollman, 2011). The meaning of a signal can be traced to rational or evolved behavioral patterns in the stage game (Lewis, 1969; Huttegger, 2007; Skyrms, 2010b).

The picture pursued here differs from the above in several respects. Firstly, we assume not just one fixed stage game and (the rationality or evolution of) a single behavioral profile for it, but an open ended class of games, called the *environment*, and (the rationality or evolution of) more general *choice mechanisms* for playing any game from that class (Fawcett, Hamblin, and Giraldeaub, 2013; McNamara, 2013; Franke and Galeazzi, 2014; Galeazzi and Franke, to appear). We are then particularly interested in how different statistical properties of the environment (the probability with which particular games are played) influences the evolution of language. Secondly, we will look at a richer picture of the agents' decision making, inspired by recent theoretically driven and empirically supported probabilistic models of language use (Frank and Goodman, 2012; Franke and Jäger, 2016; Goodman and Frank, 2016). Concretely, we assume that agents have a mental lexicon, which contains their subjective belief about lexical meanings, and that agents' language use in context is a function of pragmatic reasoning on top of this conventional meaning. We are then interested in how lexical representations evolve, when used pragmatically in a wide range of possible contexts. This is inspired by Brochhagen, Franke, and van Rooij (2017), but differs in that we here do not focus on the evolution of pragmatic reasoning strategies and instead consider not a stage game but a rich multi-game environment. Finally, while a large chunk of the game theoretic literature on meaning evolution focuses on signaling games, we here consider *reference games* (Franke, 2012b,a).

2 The environment: Multiple reference games

- The **environment** E consists of an infinite set $\mathcal{G}$ of different games, with a probability measure $P_E \in \Delta(\mathcal{G})$ such that $P_E(g)$ is the frequency or probability with which any agent would play game $g \in \mathcal{G}$.
  - The probability measure P_E will be defined in terms of a number of parameters, to be introduced in the following and summarized at the end of this section.
- We assume that all games in $\mathcal{G}$ are reference games.
  - [**alternative option**: Consider another type of game and mixtures of several game types (like in Franke (2012b)).]
- A single **reference game** is a tuple $g = \langle o_1, \ldots, o_k \rangle$ of $k \geq 2$ objects.
  - The first object o_1 is called the designated object, the designated referent or the target object/referent.
    - The speaker tries to refer to o_1; the listener guesses the referent in g; communication is successful if the listener guesses the correct referent.
  - The number k of objects varies between games.
    - E.g., an environment E could differ from an environment E' in that the expected number of objects in E is higher than that in E'. — What would that mean for an evolving language in E and E'?
- An **object** $o \in \mathcal{F}^n$ is an n-tuple of features from an n-dimensional **feature space** $\mathcal{F}^n$.
- Each **feature** $f_i \in F_i \subseteq \mathbb{R}$ of object $o = \langle f_1, \ldots, f_n \rangle$ is a value on a **property scale** F_i.
  - [**alternative option**: The property scale F_i could also include values "NA". A ball is neither rich nor poor.]
- Notice that the environment fixes, via the prior on games P_E, which feature values likely occur in which combinations. For simplicity, we can define priors from the ground up, so to speak, based on an assumption of independence of each feature dimension and objects in the environment.
  - [**alternative option**: Consider richer "stochastically structured" environments. E.g., consider cases where the several feature dimensions are correlated.]

- There are three types of property scales. They differ in the shape of the probability distribution from which feature values are sampled:

    1. open scales: $f \sim \mathcal{N}(\theta_o)$ (with θ_o a parameter; see below)
    2. half-open scales: $f \sim \text{Gamma}(\theta_h)$ (with θ_h a parameter; see below)
    3. closed scales: $f \sim \text{Beta}(1, 1)$

    - [**alternative option**: Consider categorical feature values $f \sim \text{Multinom}(\cdot)$, ordinal feature values $f \sim \text{Pois}(\cdot)$ or combinations of all the former.]
    - [**alternative option**: Consider a space of objects with fixed features (individual-level categories) and temporally variable features (stage-level categories).]

- An environment E is then defined entirely by the following parameters:

    1. λ_k — shape parameter of distribution of the size of reference games $k \sim \text{Pois}(\lambda_k) + 2$
    2. θ_o, θ_h — the shape parameters of the open and half-open property scales

- [**alternative option**: Consider reference games with only partial observations: e.g., some referents are no visible to either speaker or listener; an extreme case of this is where the speaker only observes o_1 and must reason, based on (partial) knowledge of P_E or some "heuristic", which other objects the receiver might see.]

- [**alternative option**: Consider noisy environments in which agents perceive each object only with some imprecision.]

- [**alternative option**: Compare environments which differ with respect to the properties of o_1 compared to the other objects in each g. What if o_1 is always the most/least surprising element (in some information-theoretic sense, given P_E)?]

3 Language & its users

The behavior of language users is a function of (i) their subjective representations of basic lexical meanings of words or expressions, (ii) the way in which basic messages can be combined to form expressions, and (iii) a pragmatic strategy of how to use expressions with a given semantics in a given context. The following looks at each component in turn.

3.1 Lexica & word meaning

- Agents have a **lexicon** $\mathcal{L}$ which pins down their subjective representation of the conventional/lexical meaning of content words.

- We consider the case where there is an antonym pair for every feature/property scale. E.g., if property scale F_i encodes the vertical size of an object, then there are two words, one word w_i^h for high values of F_i (think: *tall*) and another word w_i^l for low values of F_i (think: *short*).

  - [**alternative option**: The association of word meanings to feature dimensions, which is assumed to be given here, could evolve from use. This way we could also look at the evolution of category labels.]
  - [**alternative option**: The size of vocabulary could evolve, so as to see how many basic predicates are most useful and still learnable. Having a compositional system (see below) might eradicate a need for larger base-level vocabularies.]
  - [**alternative option**: Consider different systems with different vocabulary sets/sizes and ask (dryly, theoretically) which system achieves a higher communicative success on average: reason about which features to encode is most efficient (e.g., only open scales, closed scales, mixture?).]

- A lexicon $\mathcal{L}$ associates each pair $\left\langle w_i^h, w_i^l \right\rangle$ with a pair of threshold values $\mathcal{L}(w_i^{h/l}) = \theta_i^{h/l} \in F_i$. These threshold values give the boundaries above/below which w_i^h / w_i^l are true of values in F_i.

  - Intuitively, the threshold $\theta_i^{h/l}$ is the *extensional meaning* of word $w_i^{h/l}$.
    - [**alternative option**: Think of richer representations for lexical meanings; curves, intervals etc.]
    - [**alternative option**: Context dependence could be part of the lexical representation (e.g., it might matter for a threshold which class of objects a word applies to (a *short building* is likely taller than a *tall man*)). Think: lexical meaning of *tall* is the tallest $\theta_{\text{tall}} = 1/3$ of objects in a context.]

- A word w_i^h / w_i^h is true of object $o = \langle f_1, \dots, f_n \rangle$ iff $f_i \geq \theta_i^h$ / $f_i \leq \theta_i^l$.

3.2 Compositional messages

- The **language** of an agent is a set of **messages**, which is defined by a (probabilistic context-free) **grammar**.

- Each agent has his own grammar representation. Grammars need to be learned from observation and compete with each other on a measure of communicative efficiency.

- The grammar defines well-formed descriptive DPs, which we interpret as possible descriptions of the target object o_1.

- In the simplest case, which we explore first, the grammar is actually entirely trivial: it only consists of a single lexical expansion rule $DP \to w_i^{h/l}$, so that the space of possible messages is just the space of (single) words.

  - [**alternative option**: The slightly more complex case to consider next is where messages are sequences of words. (Order could matter or not, depending on whether we define incremental production and interpretation rules (see below).) Adding sequences of messages is basically adding a conjunction rule to the grammar.]
  - [**alternative option**: Add conjunction and negation to the grammar. This gives interesting stuff like *neither tall nor not tall*.]
  - [**alternative option**: Add recursive inclusion of other DPs. (Makes sense for a more structured world in which we could want to say stuff like *the tall man next to the skinny kid with the yellow shirt*).]

- The meaning of complex messages are defined as a compositional function of the lexical meanings of elements included in them and the way they are combined. (In the usual and obvious way for simple grammars: e.g. conjunction is just set-intersection.)

  - This gives us a set-theoretic interpretation of each message m in context g as $[\![m]\!]^g = \{o \in g \mid m \text{ is true of } o\}$ the set of all objects in g of which m is true.

3.3 Pragmatic language use

- We define a pair of probabilistic speaker and listener choice rules, in the spirit of the Rational Speech Act model (Frank and Goodman, 2012; Franke and Jäger, 2016; Goodman and Frank, 2016).

$$
\begin{aligned}
P_{LL}(o \mid m, g, \mathcal{L}) &= P(o \mid [\![m]\!]^g_{\mathcal{L}}) \\
P_S(m \mid o, g, \mathcal{L}) &\propto \exp(\lambda \log(P_{LL}(o \mid m, g, \mathcal{L}))) \\
P_L(o \mid m, g, \mathcal{L}) &\propto P_E(o_1 = o \mid g)\ P_S(m \mid o, g, \mathcal{L})
\end{aligned}
$$

- The speaker rule refers to a **literal listener** who interprets every message literally. Concretely, a literal listener chooses an object $o \in g$ with a probability given by Bayes rule, based on their priors (here assumed flat) and the (set-theoretic/denotational) meaning of messages.

  - [**alternative option**: Non-trivial priors for the literal listener.]

- The speaker chooses expressions based on how likely they are to make a literal listener pick the desired referent.

  - [**alternative option**: Include a cost term or utterance prior to differentiate between, e.g., simple/short and complex/long messages.]

- The pragmatic listener uses Bayes rule to reason about which referent the speaker might most likely have used.

4 Evolutionary dynamics

References

Brochhagen, Thomas, Michael Franke, and Robert van Rooij (2017). "Co-evolution of lexical meaning and pragmatic use". Manuscript Amsterdam-Tübingen.

Fawcett, Tim W., Steven Hamblin, and Luc-Alain Giraldeaub (2013). "Exposing the behavioral gambit: the evolution of learning and decision rules". In: *Behavioral Ecology* 24.1, pp. 2–11.

Frank, Michael C. and Noah D. Goodman (2012). "Predicting Pragmatic Reasoning in Language Games". In: *Science* 336.6084, p. 998.

Franke, Michael (2012a). "On Scales, Salience & Referential Language Use". In: *Amsterdam Colloquium 2011*. Ed. by Maria Aloni, Floris Roelofsen, and Katrin Schulz. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 311–320.

— (2012b). "Scales, Salience and Referential Safety: The Benefit of Communicating the Extreme". In: *The Evolution of Language: Proceedings of 9th International Conference (EvoLang 9)*. Ed. by Thomas C. Scott-Phillips et al. Singapore: World Scientific, pp. 118–124.

Franke, Michael and Paolo Galeazzi (2014). "On the Evolution of Choice Principles". In: *Proceedings of the Second Workshop Reasoning About Other Minds: Logical and Cognitive Perspectives, co-located with Advances in Modal Logic 2014, Groningen, The Netherlands, August 4th, 2014*. Ed. by Jakub Szymanik and Rineke Verbrugge. Vol. 1208. CEUR Workshop Proceedings. CEUR-WS.org, pp. 11–15.

Franke, Michael and Gerhard Jäger (2016). "Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics". In: *Zeitschrift für Sprachwissenschaft* 35.1, pp. 3–44.

Galeazzi, Paolo and Michael Franke (to appear). "Smart Representations: Rationality and Evolution in a Richer Environment". In: *Philosophy of Science*.

Goodman, Noah D. and Michael C. Frank (2016). "Pragmatic Language Interpretation as Probabilistic Inference". In: *Trends in Cognitive Sciences* 20.11, pp. 818–829.

Huttegger, Simon M. (2007). "Evolutionary Explanations of Indicatives and Imperatives". In: *Erkenntnis* 66, pp. 409–436.

Huttegger, Simon M. and Kevin J. S. Zollman (2011). "Signaling Games: Dynamics of Evolution and Learning". In: *Language, Games, and Evolution*. Ed. by Anton Benz et al. LNAI 6207. Springer, pp. 160–176.

Lewis, David (1969). *Convention. A Philosophical Study*. Cambridge, MA: Harvard University Press.

McNamara, John M. (2013). "Towards a Richer Evolutionary Game Theory". In: *Journal of The Royal Society Interface* 10.88, pp. 1–9.

Skyrms, Brian (2010a). *Signals: Evolution, Learning, and Information*. Oxford: Oxford University Press.

— (2010b). "The flow of information in signaling games". In: *Philosophical Studies* 147, pp. 155–165.