SCALES, SALIENCE AND REFERENTIAL SAFETY: THE BENEFIT OF COMMUNICATING THE EXTREME

MICHAEL FRANKE

Institute for Logic, Language & Computation
Universiteit van Amsterdam

On an abstract level, this paper tries to show how formal semantics can contribute to theories of language evolution and vice versa. More concretely, the paper argues that signaling games fail to account plausibly for a general preference to use gradable adjectives to communicate *extreme values*. The reason is that these models focus too narrowly on *descriptive language use*. Numerical simulations show that the choice of extreme values is pragmatically beneficial in situations of *referential language use* under possible noise. Bringing evolutionary modeling back to formal semantics, this yields a functional explanation for otherwise rather puzzling patterns in the use of gradable adjectives.

1. Sim-Max Games & Voronoi Languages

Lewisean signaling games (Lewis, 1969) have developed into a leading model of meaning evolution that is theoretically very well understood and has proven its value in many academic disciplines (Crawford & Sobel, 1982; Grafen, 1990; Steels, 1995; Nowak & Krakauer, 1999). More recently, a particular kind of signaling game, called *sim-max game*, has come into the focus of the linguistic community, because of its special theoretical interest to questions relating to the interface between formal semantics and pragmatics, on the one hand, and the cognitive sciences, on the other hand (Jäger & van Rooij, 2007; Jäger, 2007; Franke, Jäger, & van Rooij, 2011). A sim-max game is essentially a cheap-talk signaling game with shared preferences, where communicative success is proportional to a measure of similarity of the sender's intended meaning and the receiver's interpretation. Intended meanings and interpretations come from a metric space T with a numerical similarity measure *sim* : $T \times T \to \mathbb{R}^{\geq 0}$. One might think of this as, for example, a space of possible sense perceptions being more or less similar to one another. The sender observes a state $t \in T$, sampled from a prior distribution $\Pr(\cdot) \in \Delta(T)$, and selects a message $m \in M$. The receiver then selects an interpretation $t' \in T$. The exchange is successful proportional to the similarity of t and t', i.e., utilities for sender and receiver satisfy: $\mathrm{U}_{S,R}(t, m, t') \propto \mathit{sim}(t, t')$.

Jäger, Metzger, and Riedel (2011) prove that the evolutionarily stable states of these games give rise to so-called *Voronoi languages*. (Caveat: this short exposition cannot possibly pay due respect to all mathematical detail.) A Voronoi language is a pair $\langle \sigma, \rho \rangle$ consisting of a sender strategy $\sigma : T \rightarrow M$ and a receiver strategy $\rho : M \rightarrow T$ such that: (i) the set $\{\{t \in T \mid \sigma(t) = m\} \mid m \in M\}$ of *declarative meanings* constituted by σ is a (quasi-)partition of T into convex regions, and (ii) the *imperative meaning* $\rho(m)$ of m given ρ, is the Bayesian estimator for the cell $\{t \in T \mid \sigma(t) = m\}$. For example, if the state space is given as the unit interval $T = [0; 1]$, the prior distribution is uniform, and $\mathrm{U}(t, m, t') = -|t - t'|$, then for $|M| = 2$ there are only two isomorphic Voronoi languages which structure the state space as follows (see Figure 1): the sender will use one message for all values below ½, the other for those above; the receiver will interpret the former as ¼, the latter as ¾. For higher-dimensional spaces the resulting structure similarly is a tessellation, with T, $\mathrm{Pr}(\cdot)$ and $\mathrm{U}(\cdot)$ influencing the shapes and location of declarative and imperative meanings. In general, this technical result is conceptually highly interesting, because it demonstrates how evolutionary pressure on effective communication can explain certain natural properties of conceptual meaning space, namely the convexity of natural concepts and the centrality of prototypical interpretations (cf. Gärdenfors, 2000).

Although the Voronoi-predictions of sim-max games line up neatly also with previous results from related models on, for example, the formation of color categories (cf. Belpaeme & Bleys, 2005; Steels & Belpaeme, 2005; Baronchelli, Gong, Puglisi, & Loreto, 2010), from a linguistic point of view they are not entirely satisfactory. A case in point is the use and interpretation of *gradable adjectives*. These are adjectives like *tall*, *closed*, or *pure* which allow for comparative uses and gradual degree modification (e.g. "He is *taller* than her" as opposed to non-gradable, hence awkward "This is *more perfect* than that"). A prominent line of current semantic theorizing relates gradable adjectives to degrees on scales in a way that *prima facie* seems directly applicable to the basic set-up of sim-max games. However, it turns out that Voronoi languages do *not* capture some of the linguistically relevant subtleties in the use of gradable adjectives. This has theoretically interesting repercussions for evolutionary modeling. In order to appreciate this, a short excursion into formal semantics is necessary.

2. Degree Semantics for Gradable Adjectives

According to scale-based formal semantics of gradable adjectives the denotation of a gradable adjective A is a function $g_A : \mathrm{Dom(A)} \rightarrow D$ that maps any ap-

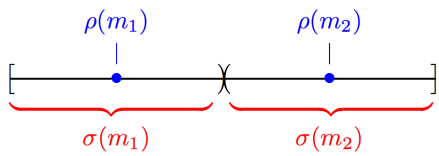

Figure 1. Simple Voronoi language on a one-dimensional state space

plicable arguments of A to a degree $d \in D$, where, crucially, $\langle D, \preceq \rangle$ is a suitably ordered *scale of degrees* (see Rotstein & Winter, 2004; Kennedy & McNally, 2005). As different adjectives may be associated with different kinds of degree scales, a simple classification scheme is obtained. Standardly, one-dimensional scales are assumed and a distinction is made as to whether these are: (i) totally open (*tall*, *short*), (ii) totally closed (*closed*, *open*), or (iii) half-open (*bent*, *pure*). Scale types explain a number of otherwise puzzling observations, such as which adjectives can combine with which modifiers. E.g., the expression *completely* A is felicitous only if A has a totally or upper-closed scale with a maximal element: compare the felicitous *completely closed* with the awkward $^{?}$*completely tall*.

Scale types also influence the licensing conditions of utterances involving gradable adjectives in positive form. Generally speaking, a simple positive sentence like "object x has property A" is considered true whenever the contextually supplied minimal degree of A-ness, $c(A)$, is no higher than $g_A(x)$. However, the contextual standard of applicability $c(A)$ is also affected by the scale type (c.f. Kennedy, 2007): if there is a $\preceq$-maximal or -minimal degree contained in $\langle D, \preceq \rangle$, then $c(A)$ is bound to this; otherwise it is to be retrieved more flexibly from the context of utterance. In more tangible terms, "*Kennedy's observation*", as we may call it here, says that adjectives which are associated with a closed scale of degrees are used rather inflexibly to denote the respective minimal or maximal values on the associated scale (modulo the usual pragmatic slack where lack of precision is conversationally harmless), whereas adjectives associated with totally open scales allow more contextual variability. In effect, this means that open-scale adjectives are more contextually variant and more prone to exhibit vagueness than closed-scale adjectives. For example, the contextual standard for the applicability of open-scale *tall* can vary considerably from one context (talking about jockeys) to another (talking about basketbal players), whereas that of closed-scale *closed* seems glued to the denotation of a minimal (zero) degree of openness.

3. The Extreme-Value Puzzle

Sim-max games fail to satisfactorily explain Kennedy's observation. The main problem, I suggest, is a more general one: sim-max games fail to explain why gradable adjectives are predominantly used to denote *extreme values*. A general preference for extreme values would explain Kennedy's observation. Suppose speakers would, for some reason or other, preferentially communicate those properties that lie at the extremes of a (one-dimensional) scale, no matter whether the absolute values of that scale may shift from context to context. In that case, it would be a mere concomitant of an extreme-value preference that closed-scale terms preferentially bind to their minimal/maximal terms (no matter what their contextually-specified actual value might be), but not so for open-scale terms. However, sim-max games fail to predict a preference for extreme values in general, and Kennedy's observation in particular, unless we make quite implausible assumptions about how to implement different scale topologies. Here is why:

Since endpoints on a one-dimensional scale have Lebesgue-measure zero, it is clear that whether T is an open, closed or half-open interval has no bearing on the structure of the resulting Voronoi language. The most natural idea is to express a difference in scale topology by a combination of T, the prior distribution $\Pr(\cdot)$ and, possibly, the utility function $\mathrm{U}(\cdot)$. Indeed, by tweaking these parameters, it is possible to set up a sim-max game in which at least the *imperative* meanings of the resulting Voronoi language take extreme values. This could be achieved by either assuming that (i) the prior distribution is convex (i.e., the more extreme a value, the more likely it is to occur), or that (ii) the utilities favor more extreme values over less extreme ones. Both assumptions are, however, inadequate. The latter assumption would stipulate exactly what ideally we want to explain, and the assumption of convex priors is simply not true in general: e.g., most people are of around average height, the chance of meeting dwarfs or giants is marginal.

It is an obvious idea to try to explain differences in usage conditions of gradable adjectives with reference to a notion of (perceptual) *salience* (e.g. Fernández, 2009). Kennedy (2007) also tries to explain the influence of scale topology on contextual usage conditions in terms of the salience of endpoints on closed scales and a principle called *Interpretive Economy* which demands that pragmatic interpretation ought to make maximal use of the available semantic resources. Since endpoints on a closed scale are salient elements of the semantic structure, these ought to be used for pragmatic interpretation. Potts (2008) rightly criticizes that *Interpretive Economy* should really be motivated by an evolutionary argument, and therefore looks at a strategic game in which speaker and hearer try to coordinate on the mutually assumed contextual standard for the use of closed-scale adjectives. Potts suggests that the endpoint convention is selected by the replicator dynamic because, by psychological salience, we may assume a (however) slight majority of extreme-value players already initially.

Salience undoubtedly plays a big role in perception and categorization. Moreover, relying on a mutual understanding of what is salient *can* increase performance, e.g., in one-shot coordination games. But not so in sim-max games where the use of extreme values would lead to strictly less effective communication. In other words, salience and the affordances of sim-max games pull in opposite directions. Consequently, just adding salience to sim-max games is *not* a sufficient explanation of an extreme value preference, because this does not explain the evolutionary *cui bono*, viz., why salient values may have adaptive linguistic value.

4. Minimal Risk of Referential Confusion

I suggest that the main reason why sim-max games fail to explain a preference for extreme values is that these games capture too narrow a notion of what language is used for. Sim-max games only model *descriptive language use*: e.g., the speaker says "John is tall" and the hearer forms a mental image of John's height. But we should also consider *referential language use*: e.g., the speaker says "John is the tall guy over there" and the hearer tries to guess who the intended referent may have been. In the latter situation, given a shared but possibly noisy context of potential referents, the speaker should ideally describe an intended referent by a property that is *minimally confusing* to the addressee. The driving intuition of this paper is that minimizing the chance of referential confusion leads to selective pressure for the communication of extreme values.

In corroboration of this intuition, consider the following simple model. Sender and receiver observe a context C that consists of n different objects. Each object is an m-tuple of values, one for each relevant property (height, thickness, …), sampled randomly from distributions corresponding to open, closed and half-open (lower-closed) scales (see the light gray lines in Figure 2 for example distributions modeling different types of scales). A context is an $n \times m$ matrix, with C_{ij} the value of property j of object i. The sender knows the *designated object* C_o that she wants to communicate, but the receiver does not. The sender selects a property j (supposing for simplicity that there is a common code for that already), and the receiver responds with choosing an object C_i. Communication would be successful iff $i = o$. But if the receiver possibly only perceives a noisy version of context C, the sender should ideally choose that property j for which the desig-

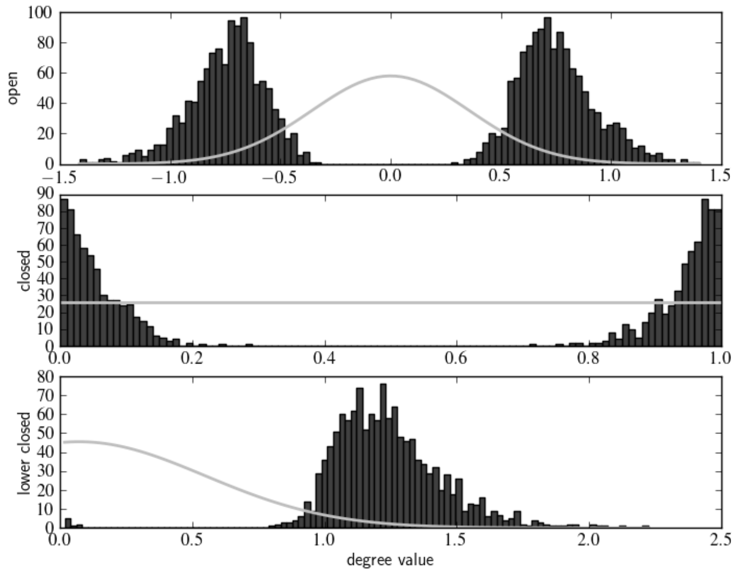

Figure 2. Frequency of values C_{oj} selected by the "optimally non-confusing" choice rule over 5000 randomly sampled contexts with $n = 30$ and $m = 24$ (8 properties each for open, closed, and half-open scales with prior distributions indicated by the light gray lines).

nated object is maximally distinct from the other objects in context. What would count as maximally distinct, and hence an optimal speaker choice, of course, depends on many different factors: noise distributions, players' knowledge thereof etc. To keep matters simple, let us assume that the probability that C_o is confused with C_i is anti-proportional to the difference $|C_{oj} - C_{ij}|$ in the respective values of the property j that is used to describe C_o. In that case, the sender's "optimally non-confusing" choice rule would be to choose arbitrarily from:

$$\arg\max_j \sum_{i=1}^{n} |C_{oj} - C_{ij}| \,. \tag{1}$$

(For clarity, this is not the only reasonably conceivable choice rule, but, perhaps, the simplest. Others have been tested with basically identical results.)

The choice rule in (1) does not yield a preference for extreme values necessarily. But if we look at the values C_{oj} that are selected by this rule for a large number of randomly sampled contexts, then indeed we see that extreme values are preferentially selected, irrespective of whether these are sampled from open, closed or half-open scales (see Figure 2). (The results in Figure 2 also indicate that our referential choice rule improves on the predictions of Voronoi languages in yet further respects: we see (distributional) *vagueness*, but also *non-complimentarity*, i.e., values which would be expressed neither by a term nor its antonym.) This suggests that a preference for extreme values, and with it Kennedy's observation, can be explained as a concomitant of the speaker's strategic choice of which property to describe a referent with in a possibly confusing environment.

5. Relating Descriptive and Referential Language Use

If correct, these considerations raise a number of technical and conceptual questions concerning the relation between descriptive and referential language use: is either one to be considered primordial? how can we study their conjoined effect on language evolution in a single model? As a partial answer to the latter issue, I suggest to look at the results of a simple but instructive model of *iterated learning* (cf. Smith, Kirby, & Brighton, 2003) in which each new generation learns its language from a finite sample of how the previous generation actually used its language. By parameterizing the proportion of descriptive/referential use, we find that descriptive and referential language use give rise to diametrically opposed selective pressures on language evolution.

The model assumes that each teacher generation is presented with contexts of objects, as above. With probability α speakers choose the property of the designated object that best matches their own (previously acquired) interpretation strategy $\rho : M \to \mathbb{R}$, as in sim-max games, and with probability $(1 - \alpha)$ speakers choose in accordance with the choice rule in (1). Learners form their interpretations $\rho(m)$ by averaging over all values expressed by the teacher generation for

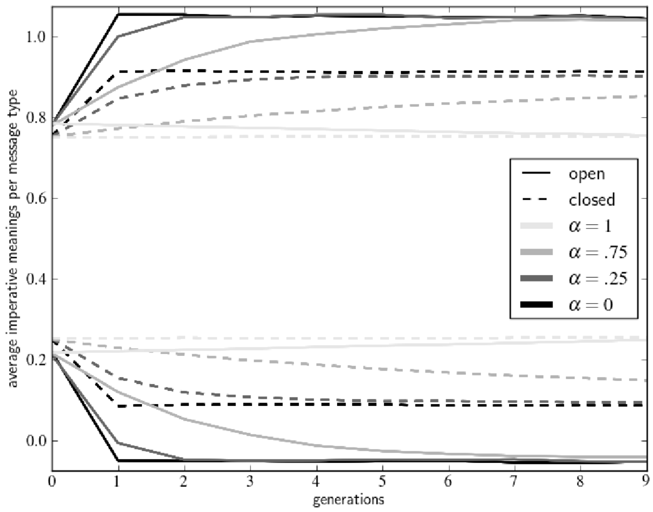

Figure 3. Temporal development of imperative meanings over 10 generations of iterated learning. Plotted are averages over 10 runs, each for different values of α. Each learner generation observed the teacher generation's output for 800 randomly sampled contexts, each containing 10 objects with 4 open- and 4 closed-scaled properties. Here, open-scale values were sampled from a normal distribution with $\mu = {}^1\!/_2$ and $\sigma = {}^1\!/_3$, closed-scale values from a uniform distribution over $[0; 1]$.

message m. (For simplicity, the initial teacher generation used a Voronoi language as interpretation strategy.) Averaged results of 20 runs of such iterated learning are plotted in Figure 3 and clearly indicate that the more prevalent referential use is, the more (and the faster) imperative meanings shift towards extreme values. For purely descriptive use ($\alpha = 1$), closed-scale terms are stuck at the Voronoi language interpretation, while open-scale interpretations even shift to less extreme values (drawn in by the higher *a priori* probability of more median values).

This shows that if the speaker has a choice which scalar property to communicate, then descriptive language use selects for less extreme, while referential language use selects for more extreme values. Hence, in order to account for a preference for extreme values (and with it Kennedy's observation), we should look rather at the affordances of referential communication, rather than descriptive communication. Though negative, this is a conceptually valuable result, in that it quite nicely charters some of the limits of signaling games as a faithful base model of linguistic communication.

In sum, I propose that taking insights from formal semantics more seriously into account, as done here, can fruitfully guide the conceptual debate and the synthetic modeling of language evolution. Formal semantics may likewise profit from evolutionary considerations, as shown here by providing a functional explanation for some otherwise perplexing patterns in the use of gradable adjectives.

References

Baronchelli, A., Gong, T., Puglisi, A., & Loreto, V. (2010). Modeling the emergence of universality in color naming patterns. *PNAS*, *107*(6), 2403–2407.

Belpaeme, T., & Bleys, J. (2005). Explaining universal color categories through a constrained acquisition process. *Adaptive Behavior*, *13*(4), 293–310.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*, 1431–1451.

Fernández, R. (2009). Salience and feature variability in definite descriptions with positive-form vague adjectives. In *Proceedings of pre-cogsci.*

Franke, M., Jäger, G., & van Rooij, R. (2011). Vagueness, signaling & bounded rationality. In T. Onoda, D. Bekki, & E. McCready (Eds.), *Jsai-isai 2010* (pp. 45–59). Springer.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought.* MIT Press.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*, 517–546.

Jäger, G. (2007). The evolution of convex categories. *Linguistics and Philosophy*, *30*(5), 551–564.

Jäger, G., Metzger, L. P., & Riedel, F. (2011). *Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals.* (To appear in *Games and Economic Behavior*)

Jäger, G., & van Rooij, R. (2007). Language structure: Psychological and social constraints. *Synthese*, *159*(1), 99–130.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*, 1–45.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, *81*(2), 345–381.

Lewis, D. (1969). *Convention. a philosophical study.* Harvard University Press.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *PNAS*, *96*, 8028–8033.

Potts, C. (2008). *Interpretive Economy, Schelling Points, and evolutionary stability.* (Manuscript, UMass Amherst)

Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics*, *12*(3), 259–288.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, *9*, 371–386.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, *2*(3), 319–332.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for color. *Behavioral and Brain Sciences*, *28*(4), 469–529.