Evolutionary motivations for semantic universals

Robert van Rooij*

Abstract

Most work in 'evolutionary linguistics' seeks to motivate the emergence of linguistic universals. Although the search for universals never played a major role in semantics, a number of such universals have been proposed concerning connectives, property and preposition denoting expressions, and quantifiers. In this paper we suggest some evolutionary motivations for these proposed universals using game theory.

1 Introduction

The majority of work on the evolution of language concentrates on the evolution of syntactic and phonetic rules and/or principles, and says virtually nothing about semantics. This is, on the one hand, surprising, because languages without meanings associated with their expressions hardly make any sense. This is obvious if language is thought of as the main vehicle to transmit information, but is equally true if one thinks that natural languages are primarily used as internal representation codes in which thinking can be carried out. On the other hand, however, the focus on syntactic principles is understandable, because syntax seems to give the evolutionary linguist more things to explain. This is due to the fact that in contrast to semantics, in syntax the search for *universals* always played an important role. The major goal of linguists working in the Chomskian tradition is to find the grammatical principles that isolate the subclass of all possible *human* languages from the class of all

*This paper was presented at the Blankensee conference in Berlin and at the first Scottish-Dutch workshop on Language Evolution, both in the summer in 2005. I would like to thank the organizers of those workshops and the participants for their comments. In particular, I would like to thank Bernhard Schröder for discussions on section 3.2 of this paper, and Gerhard Jäger, Samson de Jager, and an anonymous reviewer for their comments on an earlier version of this paper.

possible languages. This set of abstract grammatical principles forms then the universal grammar, and it is exactly the features of this universal grammar, or language faculty, that most evolutionary linguists want to explain. The search for universals as innate features of the language faculty played traditionally a much less important role in semantics. In traditional typological work, the discussion of semantic universals is normally limited to color- and kinship terms (cf. Leech, 1974). This lack of interest in universals can partly be explained by the fact that denotational semantics – the standard, and certainly most productive approach towards natural language meaning – stems from logicians like Frege and Montague who held a rather anti-psychologistic view towards their subject matter. More recently, however, the search for universals started to play a more important role here as well. One of the reasons for this was that around that time semanticists became more aware of the fact that although making use of sets and functions of many different types allows one to describe the meanings of expressions in a simple and compositional way, an unlimited use of this machinery is all too powerful, and would demand too much of our cognitive resourses.

There are, in fact, many constraints involving the interpretation of their expressions, or at least in their use, shared by all languages of the world. For instance, it seems to be the case that in all languages more can be communicated than what is explicitly said, e.g., all languages make use of conversational *implicatures*. The exact mechanisms by which we are allowed to do so are still controversial, but it is safe to assume that the explanation for this fact of language (use) involves considerations of efficiency. Similarly, in terms of utility and efficiency one can explain why the use of negative predicates is marked, although logically there is no reason for this. Other semantico/pragmatic universals concern also linguistic *structure*. One can observe, for instance, that of all the speech acts that we can express in natural language, only three of them are normally grammaticalized, and distinguished, in mood (i.e., declarative, imperative, and interrogative). Finally there are universals that make claims about what kinds of meanings are expressed by short and simple terms in natural languages. One of them concerns *indexicals*, short expressions corresponding to the English *I*, *you*, *this*, *that*, *here*, etc., the denotations of which are essentially context-dependent. It seems that all languages have short words that express such meanings (cf. Goddard, 2001), and this fact makes evolutionary sense: it is a useful feature of a language if it can refer to nearby individuals, objects, and places, and we can do so by using short expressions because their denotations can normally be inferred from the shared context between speaker and hearer. The latter kind of semantic universal is what I am most interested in in this paper: of all possible meanings that we *can* potentially express in natural language, which ones are expressed in 'simple', or lexicalized, terms

in all languages? The major goal of this paper, however, is not to come up and state such universals. Rather, I will rely on some descriptive work on this area and concentrate myself on giving evolutionary motivations of these universals. While admitting that Chomskian-like 'explanations' by innateness might be used as a last resort, we would like to find deeper motivations for why we have simple expressions for some particular meanings but not for others in natural language. Most naturally, our explanatory motivation should make use of notions like *utility*, *learnability* and *complexity*: we typically want to express those meanings in simple terms that are (i) useful, (ii) easy to learn and remember, and (iii) easy to use.

As a formal model of evolution, I will make use of evolutionary game theory (EGT). This theory stems from biology to model *natural* selection, but has acquired an important role in economics as well to model *cultural evolution*. As a consequence – and perhaps in contrast to what is suggested by Nowak and associates – an 'explanation' of universal features of natural language in terms of EGT is quite neutral on the issue of whether these features have biologically evolved to become part of our 'language faculty' or that they are established anew by reinforcement due to language learning or use.[1]

2 Evolution and signaling games

David Lewis (1969) introduced *signaling games* to account for linguistic conventions, and these games were developed further in economics and theoretical biology. In this framework, signals have an underspecified meaning, and the actual interpretation the signals receive depend on the equilibria of sender and receiver strategy combinations of such games. Recently, these games have been looked upon from an *evolutionary* point of view to study the evolution of language. I will first sketch these games here and then look at them from an evolutionary point of view.

A signalling game is a two-player game with a *sender*, s, and a *receiver*, r. This is a game of *asymmetric* information: The sender starts off knowing something that the receiver does not know. The sender knows the state $t \in T$ she is in but has no substantive payoff-relevant actions.[2] The receiver has a range of payoff-relevant actions to choose from but has no private information, and his prior beliefs concerning the state the sender is in are given by a probability distribution P over T; these prior beliefs are common knowledge. The sender, knowing t and trying to influence the action of the receiver, sends

[1] See Jäger & van Rooij (to appear) for extensive discussion on how to most naturally interpret EGT to account for these universal features.

[2] In game theory, it is standard to say that t is the *type* of the sender.

to the latter a signal of a certain message m drawn from some set M. The messages don't have a pre-existing meaning. The other player receives this signal, and then takes an action a drawn from a set $\mathcal{A}$. This ends the game. Notice that the game is *sequential* in nature in the sense that the players don't move simultaneously: the action of the receiver might *depend* on the signal he received from the sender. For simplicity, we take T, M and $\mathcal{A}$ all to be finite. A pure *sender strategy*, S, says which message the sender will send in each state, and is modeled as a (deterministic) *function* from states to signals (messages): $S \in [T \to M]$. A pure *receiver strategy*, R, says which action the receiver will perform after he received a message, and is modeled as a (deterministic) function from signals to actions: $R \in [M \to \mathcal{A}]$. Mixed strategies (probabilistic functions, which allow us to account for ambiguity) will play only a minor role in this paper and can for the most part be ignored.

As an example, consider the following signalling game with two equally likely states: t_1 and t_2; two signals that the sender can use: m_1 and m_2; and two actions that the receiver can perform: a_1 and a_2. Sender and receiver each have now four (pure) strategies:

Sender :

	t_1	t_2
S_1	m_1	m_2
S_2	m_1	m_1
S_3	m_2	m_1
S_4	m_2	m_2

Receiver :

	m_1	m_2
R_1	a_1	a_2
R_2	a_2	a_1
R_3	a_1	a_1
R_4	a_2	a_2

Sender strategy S_1, for instance, says that s sends message m_1 in state t_1 and message m_2 in state t_2, while receiver strategy R_3 reacts by action a_1 independent on which message he receives.

To complete the description of the game, we have to give the *payoffs*. The payoffs of the sender and the receiver are given by utility functions U_s and U_r, respectively, which state for each state-action pair its payoff, modeled by a real number.[3] Formally, they are functions from $T \times \mathcal{A}$ to the set of reals, $\mathbf{R}$.

Coming back to our example, we can assume, for instance, that the utilities of sender and receiver are in perfect alignment – i.e., for each agent i, $U_i(t_1, a_1) = 1 > 0 = U_i(t_1, a_2)$ and $U_i(t_2, a_2) = 1 > 0 = U_i(t_2, a_1)$:[4]

$U_i(t, a)$	a_1	a_2
t_1	1	0
t_2	0	1

[3]Just like Lewis (1969) we assume that sending messages is costless, which means that we are talking about *cheap talk* games here.

[4]This assumption allows Hurford (1989), Nowak & Krakauer (1999) and others to represent sender and receiver strategies by convenient transmission and reception matrices.

Notice that according to this utility function, action a_1 is for both the preferred action in state t_1, while action a_2 is for both the best in state t_2. We can model this by means of a 1-1 function f from situations to actions: $f(t_1) = a_1$ and $f(t_2) = a_2$. This 1-1 function will play an important part later in the paper.

An *equilibrium* of a signalling game is described in terms of the strategies of both players. If the sender uses strategy S and the receiver strategy R, it is clear how to determine the utility of this profile for the sender, $U_s^*(t, S, R)$, in any state t:

$$U_s^*(t, S, R) \quad = \quad U_s(t, R(S(t)))$$

The receiver does not know in which situation he is, which makes things a bit more complicated for him. Because it might be that the sender using strategy S sends in different states the same signal, m, the receiver doesn't necessarily know the unique state relevant to determine his utilities. Therefore, he determines his utilities, or *expected* utilities, with respect to the *set* of states in which the speaker could have sent message m. Let us define S_t to be the *information state* (or information set) the receiver is in after the sender, using strategy S, sends her signal in state t, i.e. $S_t = \{t' \in T : S(t') = S(t)\}$. With respect to this set, we can determine the (expected) utility of receiver strategy R in information state S_t, which is R's expected utility in state t when the sender uses strategy S, $U_r^*(t, S, R)$ (where $P(t'|S_t)$ is the conditional probability of t' given S_t):

$$U_r^*(t, S, R) \quad = \quad \sum_{t' \in T} P(t'|S_t) \times U_r(t', R(S(t')))$$

A strategy profile $\langle S, R\rangle$ forms a *Nash equilibrium* iff neither the sender nor the receiver can do better by unilateral deviation. That is, $\langle S, R\rangle$ forms a Nash equilibrium iff for all $t \in T$ the following two conditions are obeyed:

$$\begin{array}{ll} \text{(i)} & \neg\exists S' : U_s^*(t, S, R) < U_s^*(t, S', R), \text{ and} \\ \text{(ii)} & \neg\exists R' : U_r^*(t, S, R) < U_r^*(t, S, R'). \end{array}$$

As can be checked easily, our game has 6 Nash equilibria: $\{\langle S_1, R_1\rangle, \langle S_3, R_2\rangle, \langle S_2, R_3\rangle, \langle S_2, R_4\rangle, \langle S_4, R_3\rangle, \langle S_4, R_4\rangle\}$. This set of equilibria depends on the receiver's probability function. If, for instance, $P(t_1) > P(t_2)$, then $\langle S_2, R_4\rangle$ and $\langle S_4, R_4\rangle$ are no equilibria anymore: it is always better for the receiver to perform a_1.

In signalling games it is assumed that the messages have no pre-existing meaning. However, it is possible that meanings can be associated with them due to the sending and receiving strategies chosen in equilibrium. If in equilibrium the sender sends different messages in different states and also the

receiver acts differently on different messages, we can say with Lewis (1969, p. 147) that the equilibrium pair $\langle S, R\rangle$ fixes meaning of expressions in the following way: for each state t, the meaning of message $S(t)$ can either be thought of *descriptively* as S_t as the *set of states* in which message $S(t)$ is sent: $S_t = \{t' \in T | S(t') = S(t)\}$, or *imperatively* as the *action* performed by the receiver: $R(S(t))$. Notice that S_t is just the same as $S^{-1}(S(t))$ – the inverse sender strategy applied to message $S(t)$ – a notion that will play an important role in section 3 of this paper. Following standard terminology in economics (e.g. Crawford & Sobel, 1982), let us call $\langle S, R\rangle$ a (fully) *separating* equilibrium if there is a one-to-one correspondence between states (meanings) and messages, i.e., if there exists a bijection between T and M. Notice that among the equilibria in our example, two of them are separating: $\langle S_1, R_1\rangle$ and $\langle S_3, R_2\rangle$. According to Lewis (1969), only these separating equilibria are appropriate candidates for being a convention, and he calls them *signaling systems*.

Unfortunately, however, Lewis's characterization of conventions as separating equilibria has a number of difficulties. First, Lewis is working in the framework of traditional, or rational, game theory, where unreasonably strong assumptions have to be made concerning rationality and common knowledge in order to motivate the equilibria from an individualistic point of view. Second, not all Nash equilibrium languages of signaling games are separating, so, Lewis cannot really give a game-theoretical motivation for why linguistic conventions – as being separating equilibria – are more self-purporting than equilibria that are not separating.

As it turns out, if we look upon signaling games from an evolutionary, rather than a standard rationalistic point of view, both of Lewis's problems can be solved. Thinking of signaling games from an evolutionary perspective, it is natural to think of players as *users of languages*, rather than as senders or receivers. To do this, it is useful to think of signaling games first from a strategic perspective.

In the above description of a Nash equilibrium, we have taken the game to be one in *extensive form*, where the sender acts first, and has information that the receiver lacks. But we can also think of the game from a *strategic* point of view, according to which these asymmetries of action and information no longer holds. If we think of the game from a strategic point of view, we assume that also the sender has incomplete information about the actual state, and this information is represented by a probability function. We assume that the information sender and receiver have is the same, and can be represented by a *a common* (prior) probability distribution P. Sender and receiver now have to make their choices simultaneously, and these depend on their expected utilities. These expected utilities are defined in terms of the common probabil-

ity distribution P: $EU_i(S,R) = \sum_{t \in T} P(t) \times U_i(t, R(S(t)))$, with $i \in \{s, r\}$. A combination of a sender and receiver strategy $\langle S, R\rangle$ is now a Nash equilibrium if neither the sender nor the receiver can receive a higher expected utility by unilateral deviation. Thus $\langle S, R\rangle$ is a Nash equilibrium iff the following two conditions are met:

$$\begin{array}{ll} \text{(i)} & \neg\exists S' : EU_s(S,R) < UE_r(S',R), \text{ and} \\ \text{(ii)} & \neg\exists R' : EU_r(S,R) < EU_r(S,R'). \end{array}$$

Until now we have assumed implicitly that individuals have fixed roles in coordination situations: they are always either a sender or a receiver. In this sense it is an *asymmetric* game. It is natural, however, to give up this assumption and turn it into a *symmetric* game: we postulate that individuals both speak and listen, and can take both the sender- and the receiver-role. Now we might think of strategies as that what individuals will do if they play their two roles. An individual strategy can now be modeled as a pair like $\langle S, R\rangle$ and can be thought of as a *language*. Notice that because in our example we had 4 sender strategies and 4 receiver strategies, there will now be 16 individual strategies, or languages, that individuals can choose between. Defining S_i and R_i as the sender and receiver strategies of language L_i, we take $U_s(t, L_i, L_j) = U(t, R_j(S_i(t)))$ and $U_r(t, L_i, L_j) = U(t, R_i(S_j(t)))$.

Consider now the symmetric strategic game in which each player can choose between languages. Notice that for our very simple example this is already a 16×16 game, which is hard to represent by a payoff table on one page. To determine the payoffs of the (not represented) table, we have to know how to define the utility of playing language L_i with an other agent who plays L_j. We assume that this will be the same as the utility of playing L_j with an other agent who plays L_i. On the assumption that individuals take both the sender and the receiver role half of the time, the following utility function, $\mathcal{U}(L_i, L_j)$, is natural for an agent with strategy L_i who plays with an agent using L_j (where $EU_i(L, L')$ denotes the expected utility for i to play language L if the other participant plays L', i.e. $\sum_t P(t) \times U_i(t, L, L')$):

$$\begin{array}{rcl} \mathcal{U}(L_i, L_j) & = & [\frac{1}{2} \times (\sum_t P(t) \times U_s(t, L_i, L_j))] + [\frac{1}{2} \times (\sum_t P(t) \times U_r(t, L_i, L_j))] \\ & = & \frac{1}{2} \times (EU_s(L_i, L_j) + EU_r(L_i, L_j)). \end{array}$$

Suppose now that two players are playing a language game. The pair $\langle L_i, L_j\rangle$ is a Nash equilibrium under the standard condition that no player can profit from unilateral deviation. Because the language game is symmetric (meaning that $\mathcal{U}_1(L_i, L_j) = \mathcal{U}_2(L_j, L_i)$), we are interested in so-called symmetric equilibria where each agent chooses the same strategy, i.e., that the language chosen is a *convention*. Then we say that L_i is a symmetric (Nash)

equilibrium of the language game iff L_i is an optimal language to use if the other player uses L_i as well. Thus L_i is a symmetric Nash equilibrium iff $\mathcal{U}(L_i, L_i) \geq \mathcal{U}(L_j, L_i)$ for all languages L_j. It is straightforward to show that language L_i is a (strict) symmetric equilibrium of the (symmetric) language game if and only if the strategy pair $\langle S_i, R_i \rangle$ is a (strict) equilibrium of the (asymmetric) signalling game. But this means that our move from asymmetric signaling games to symmetric language games does not solve Lewis's (1969) problems mentioned above. These problems can be solved, however, if we think of language games from an evolutionary point of view and look at evolutionarily stable states rather than at Nash equilibria.

The first problem is that the rationality and knowledge assumptions required to explain the attractiveness of a Nash equilibrium in standard game theory are unreasonably strong. Evolutionary game theory (EGT) doesn't make such strong assumptions, but rather takes individuals as very limited in their computational resources in their ability to reason and takes them to be fully uninformed about what strategies other individuals will choose. The notion of stability used in EGT is not reached in one go by thorough interactive reasoning of the agents, but reached after a (possibly) long trajectory of trial and error by all agents involved. Thus, using evolutionary rather than standard game theory to explain linguistic conventions as equilibria seems to solve Lewis's first problem. Lewis's second problem was to give a motivation for the selection of separating equilibria among other equilibria. As we will see soon, the separating equilibria correspond exactly with the languages that are evolutionarily stable in our simple setup.

So, under what circumstances is language L evolutionarily stable in our language game? Thinking of strategies immediately as languages, standard evolutionary game theory (see e.g. Maynard Smith, 1982) gives the following answer. Suppose that all individuals of a population use language L, except for a small fraction ϵ of 'mutants' which have chosen language L'. Assuming random pairing of strategies, the expected utility, or *fitness*, of language $L_i \in \{L, L'\}$, $\mathcal{EU}^{\epsilon}(L_i)$, is now:

$$\mathcal{EU}^{\epsilon}(L_i) = (1 - \epsilon)\mathcal{U}(L_i, L) + \epsilon\mathcal{U}(L_i, L')$$

In order for mutation L' to be driven out of the population, the expected utility of the mutant need to be less than the expected utility of L, i.e., $\mathcal{EU}^{\epsilon}(L) > \mathcal{EU}^{\epsilon}(L')$. But this means that either $\mathcal{U}(L, L)$ is higher than $\mathcal{U}(L', L)$, or these are the same but $\mathcal{U}(L, L')$ is higher than $\mathcal{U}(L', L')$. Thus, we have derived Maynard Smith & Price's (1973) definition of an evolutionarily stable strategy (ESS) for our language game.

Definition 1 *(Evolutionarily Stable Strategy, ESS)*

Language L is Evolutionarily Stable in the language game with respect to mutations if

1. $\mathcal{U}(L, L) \geq \mathcal{U}(L', L)$ *for every* L', *and*
2. *if* $\mathcal{U}(L, L) = \mathcal{U}(L', L)$, *then* $\mathcal{U}(L', L') < \mathcal{U}(L, L')$.

Because the first condition means that $\langle L, L \rangle$ is a Nash equilibrium, we see that the standard equilibrium concept in evolutionary game theory is a *refinement* of its counterpart in standard game theory. As it turns out, this refinement gives us an alternative way from Lewis (1969) to characterize the Nash equilibria that are good candidates for being a convention.

In an interesting article, Wärneryd (1993) proves the following result: For any sender-receiver game of the kind introduced above, with the same number of signals as states and actions, a language $\langle S, R \rangle$ is evolutionarily stable if and only if it is a (fully) separating Nash equilibrium.[5] But this means that our evolutionary perspective upon signaling games explains why only languages that give a 1-1 mapping between signals and meanings (whether imperative or descriptive) can be evolutionarily stable. Such a language is, of course, a very simple holistic language, but in this paper we will discuss to what extent signaling games from an evolutionary perspective can be used to discuss more interesting features of languages. Wärneryd's result is interesting, but not enough for our purposes: it could be that stable states can never be reached. To see whether they can be reached, we have to look at the dynamics 'behind' EGT.

Taylor & Jonker (1978) defined their *replicator dynamics* to provide a continuous dynamics for evolutionary game theory. It tells us how the distribution of strategies playing against each other changes over time. For our language game this can be done as follows: On the assumption of random pairing, the *expected utility*, or fitness, of language L_i at time t, $\mathcal{EU}_t(L_i)$, is defined as:

$$\mathcal{EU}_t(L_i) = \sum_j P_t(L_j) \times \mathcal{U}(L_i, L_j).$$

where $P_t(L_j)$ denotes the proportion of individuals in the population at time t that play language L_j. The expected, or average, utility of a population of

[5] This result doesn't hold anymore when there are more signals than states (and actions). We will have some combinations $\langle S, R_i \rangle$ and $\langle S, R_j \rangle$ which in equilibrium give rise to the same behavior, and thus payoff, although there will be an unused message m where $R_i(m) \neq R_j(m)$. Now these combinations are separating though not ESS. Wärneryd defines a more general (and weaker) evolutionary stability concept, that of an evolutionarily stable *set*, and shows that a strategy combination is separating if and only if it is an element of such a set.

languages $\mathbf{L}$ with probability distribution P_t is then:

$$\mathcal{EU}_t(\mathbf{L}) = \sum_{L \in \mathbf{L}} P_t(L) \times \mathcal{EU}_t(L).$$

The *replicator dynamics* (for our language game) is then defined as follows:

$$\frac{\mathrm{d}P(L)}{\mathrm{d}t} = P(L) \times (\mathcal{EU}(L) - \mathcal{EU}(\mathbf{L})).$$

A *dynamic equilibrium* is a fixed point of the dynamics under consideration. In a language game with 16 languages, the vector $\langle P(L_1), ..., P(L_{16})\rangle$ is a fixed point of the language game dynamics if for each language L_i, $\frac{\mathrm{d}P(L_i)}{\mathrm{d}t} = 0$. This means that the proportion of languages doesn't change anymore over time once such a probability distribution is reached. A dynamic equilibrium is said to be *asymptotically stable* if (intuitively) a solution path where a small fraction of the population starts playing a mutant strategy still converges to the stable point. Asymptotic stability is a refinement of the Nash equilibrium concept. And one that is closely related with the concept of ESS. Taylor & Jonker (1978) show that every ESS is asymptotically stable. Although in general it isn't the case that all asymptotically stable strategies are ESS, on our assumption that a language game is a *cooperative* game (and thus *doubly* symmetric)[6] this is the case. Thus, we have the following

Fact 1 *A language L is an ESS in our language game if and only if it is asymptotically stable in the replicator dynamics.*

This result is appealing, but not exactly enough for our purposes. What we would like to see is that a separating Nash equilibrium will evolve in our evolutionary language game *by necessity*. But the above result does not imply that every solution will converge toward an asymptotically stable state. In fact, as shown independently by Huttegger (to appear) and Pawlowitsch (ms), our evolutionary language games have stable (mixed) states (or better, neutrally stable states) which are not asymptotically stable. However, Huttegger (p.c.) claims that if we add a little bit of mutation to the evolutionary dynamics, all stable states will be asymptotically stable, from which we *can* infer our desired conclusion.

In general, it is very hard to find out what (asymptotic) fixed points look like. For simple 2×2 symmetric games, however, this is relatively easy. Suppose we have a symmetric language game with only two languages involved,

[6]Our symmetric language games are *doubly* symmetric because for all $L_i, L_j, \mathcal{U}(L_i, L_j) = \mathcal{U}(L_j, L_i)$.

with the following payoff table (with $a - d \geq 0$ and $b - c \geq 0$):

$\mathcal{U}(L_i, L_j)$	L_1	L_2
L_1	a, a	c, d
L_2	d, c	b, b

Suppose that at time t the proportion of individuals playing language L_1 is $P_t(L_1)$. The replicator dynamics for L_1 says that this proportion will be $P_t(L_1) \times (\mathcal{EU}_t(L_1) - \mathcal{EU}_t(\mathbf{L}))$ in the next time-point t'. But this is just

$$P_{t'} = P_t(L_1) \times ((P_t(L_1) \times a) + P_t(L_2) \times c) - [[P_t(L_1) \times (P_t(L_1) \times a) + P_t(L_2) \times c))] \\ +[P_t(L_2) \times ((P_t(L_1) \times d) + P_t(L_2) \times b))]].$$

Let us now abreviate $P_t(L_1)$ by p, $P_t(L_2)$ thus by $(1 - p)$, and $P_{t'}(L_1)$ by p'. Then the last formula abbreviates to

$$p' = p\times((p\times a)+((1-p)\times c))-[[p\times(p\times a)+((1-p)\times c))]+[((1-p)\times((p\times d)+(1-p)\times b))]].$$

This simplifies to

$$p' = p \times [(1-p) \times [(p \times a) + ((1-p) \times c)]] - [(1-p) \times [(p \times d) + ((1-p) \times b)]].$$

This again simplifies to

$$p' = p \times (1-p) \times [(p \times a) + ((1-p) \times c)) - (p \times d) + ((1-p) \times b)],$$

which finally results in

$$p' = p \times (1-p) \times [p \times (a-d) + ((1-p) \times (c-b))].$$

The vector of probability distributions $\langle p, (1-p)\rangle$ is a restpoint of the replicator dynamics iff $p = p'$ (and thus $(1-p) = (1-p')$). This is obviously the case iff one of the three arguments of the above formula is 0. Thus this is the case if either (i) $p = 0$, (ii) $(1-p) = 0$ and thus $p = 1$, or (iii) $[p \times (a-d) + ((1-p) \times (c-b))] = 0$. The latter is the case whenever:

$$\begin{array}{lll}
(p \times (a-d)) + ((1-p) \times (c-b)) & = 0 & \text{iff} \\
p \times (a-d) + (c-b) - (p \times (c-b)) & = 0 & \text{iff} \\
(c-b) + (p \times (a-d)) - (p \times (c-b)) & = 0 & \text{iff} \\
(p \times (c-b)) - (p \times (a-d)) & = c-b & \text{iff} \\
(p \times c) - (p \times b) - (p \times a) + (p \times d) & = c-b & \text{iff} \\
(p \times (-(a+b) + (c+d)) & = c-b & \text{iff} \\
\frac{(c-b)}{-(a+b)+(c+d)} & = p. &
\end{array}$$

Although the vector $\langle p, (1-p)\rangle$ is a restpoint of the replicator dynamics if $p = \frac{(c-b)}{-(a+b)+(c+d)}$, this restpoint is *not asymptotically stable*. If p is any higher than $\frac{(c-b)}{-(a+b)+(c+d)}$, the proportion of individuals in the population that play language L_1 will grow, and we will eventually end up with a population where all individuals play this language. If p is lower than the above fraction, we will end up eventually with a population where all individuals play L_2. Thus only situations where either everybody plays L_1 or everybody plays L_2 are asymptotically stable. The condition under which one of the languages will grow will play a role in section 3.2 of this paper.

3 Towards natural language

As we have seen above, evolutionary game theory predicts that, if there are equally many messages as there are situations, in all and only all evolutionarily stable languages there exists a 1-1-1 correspondence between situations, messages, and actions, if in each situation there is exactly one action that is optimal for both speaker and hearer. It is obvious that in this simple communication system there can be no role for property-denoting expressions and connectives: the existence of a property-denoting message, or of a disjunctive or conjunctive message, would destroy the 1-1 correspondence between (types of) situations and messages. That gives rise to the question, however, under which circumstances messages with such more complex meanings could arise.

3.1 Properties

As noted in the introduction, standard formal semantics predicts that (for a simple fragment) any function from the set of worlds to the set of functions from individuals to truth values is a property that can be denoted by a property denoting expression. It is obvious, however, that in any language only a tiny fragment of all these functions are, in fact, denoted by simple words or constructions. This gives rise to the following questions: (i) can we *characterize* the properties that are denoted by simple expressions in natural language(s), and, if so, (ii) can we give a pragmatic and/or evolutionary *explanation* of this characterization?

The first idea to limit the use of all possible properties that comes to mind, is that only those properties will be expressed a lot in natural language that are *useful* for sender and receiver. Using our signaling game framework, it is easy enough to show how usefulness can influence the existence of property denoting terms when we either have less messages, or less actions than we have

situations.[7]

Let us first look at the circumstances under which the signaling strategy sends the same message in equilibrium in different situations, when there are less signals than situations. Suppose that we have three sitations, three actions, but only two messages. Because the receiver strategy is a function from messages to actions, in equilibrium there can only be two actions really be performed. Which of those actions that will be depends on the utilities and probabilities involved. Consider the following utility tables:

$U(t,a)$	a_1	a_2	a_3
t_1	6	0	0
t_2	0	4	0
t_3	0	1	2

$U(t,a)$	a_1	a_2	a_3
t_1	4	0	0
t_2	0	4	0
t_3	0	1	4

In both cases there exists a 1-1 correspondence between situations and messages. If there are three messages, in each situation the sender will send a different message, and the receiver will react appropriately. When there are only two messages, however, expected utility will play a role. In the left-hand table above it is more useful to distinguish t_1 from t_2 and t_3, then to distinguish t_2 from t_3. As a consequence, in equilibrium t_2 and t_3 will not be distinguished from each other and in both situations the same message will be sent (the receiver will then perform action a_2). We have implicitly assumed here that the probability of the three situations was equal. Consider now the table on the right-hand side, and suppose that the probability that t_1 is the case is $\frac{5}{9}$, $P(t_1) = \frac{5}{9}$, while $P(t_2) = \frac{3}{9}$ and $P(t_1) = \frac{1}{9}$. Again, it will be more useful to distinguish t_1 from t_2 and t_3, then to distinguish t_2 from t_3. Thus, also here we find that in equilibrium t_3 will not be distinguished separately, and meshed together with t_2.[8]

Utility also plays an important role when there are less (relevantly different) actions than situations. Consider the well-known alarm call signaling system of vervet monkeys: what has evolved is a signaling system in which 3 different predators (snake, eagle, and leopard) are correlated with 3 different signals. This signaling system made sense because in the three (relevantly) different situations, three different actions were triggered by the fellow vervet monkeys:

[7]These abstract formulations might be used to model other 'real-world' phenomena as well, such as noise in the communication channel which doesn't allow receivers to discriminate enough signals; a limitation of the objects speakers are acquainted with, perhaps due to ever changing contexts; and maybe also non-aligned preferences between sender and receiver.

[8]It should be noted, though, that in case $U(t_3, a_2)$ were 0, the clustering where t_3 is meshed with t_1 would be equally good in both cases. Thus, the clustering that emerges can depend crucially on the specific numbers in the table. See Donaldson et al (ms) for more discussion.

look on the ground, look upward, and go to the nearest tree, respectively. Suppose now, however, that for two of these different predators (say snake and leopard) the same action was called for. In that case, the vervet monkeys have no 'reason' to send a different alarm call when a snake or when a leopard approaches. In fact, we can now think of the situation as having only two (relevantly) different states, and two actions. In equilibrium, two messages are used; one for each relevantly different state, where the meaning of 'relevantly different' depends on which action should be performed in that state.[9] This discussion is obviously relevant for the emergence of a language with property denoting expressions: although individuals might be able to distinguish among several different situations, or objects, there is no reason to do this, as reflected in language, when it doesn't have any practical use. Thus, some messages will denote in equilibrium sets of states, not because the agents cannot distinguish between the states, but because making a distinction is not useful.[10] Putting this in a slightly different way: in natural language we collect different objects together to form a property, when there is a practical incentive to scramble them together.

A common complaint of Chomskian linguists (e.g. Bickerton, Jackendoff) against explanations like the one above is that usefulness can't be the only constraint: there are many useful properties, or distinctions 'out there' that are still not really named, or distinguished, in simple natural language terms. It seems that other, additional, constraints are called for, constraints that should be explained in terms of *learnability* and/or *complexity*. It is easy to see that a property is easier to learn, remember, or use, when it has some structural features. Bickerton (1981), for instance, hypothesizes that 'simple' expressions can only denote *connected*, or *convex*, regions of cognitive space, and hypothesizes that the preference for convex properties is an innate property of our brains. Unfortunately, when we think of properties as in standard

[9] I have been somewhat sloppy here. If these monkeys still have three or more signals at their disposal, an evolutionarily stable strategy is ruled out for technical reasons: there will be different strategy combinations that give rise to the same 1-1 correspondence between situations and actions, but differ only in the unused message. Because such strategy combinations give rise to the same behavior, and do equally well, none of them can be an evolutionary stable strategy, ESS. However, these strategy combinations will all be part of a set of neutrally stable strategies (Wärneryd, 1993), a slightly weaker notion of stability than ESS, but one that reduces to it in case there are equally many states, messages, and actions.

[10] There are other reasons why, in equilibrium, a non-separating language will evolve. One reason might be that there is a difference in preferences between the agents on the actions to be performed in the situations. Another circumstance, discussed in Nowak & Krakauer (1999), is when there is noise in the signaling system: the hearer is not sure which signal is being sent. In this paper I will limit myself to fully cooperative games, however, with noiseless channels.

denotational semantics, it is impossible to distinguish properties that have, from properties that don't have such 'structural features'. Partly for reasons like this both linguists and philosophers (van Fraassen 1967, Stalnaker, 1981) proposed alternative, and *richer*, frameworks to represent meanings and suggested that they can be used to distinguish 'natural' from 'unnatural' properties. A prominent idea found in those proposals is to assume that meaning spaces have spatial, or *geometrical* structure, and should be modeled as *vector spaces*. Properties are now just subsets of this vector, or meaning space. All properties distinguished within standard denotational semantics can be distinguished in terms of these meaning spaces as well. However, because meaning spaces have some additional structure, i.e., the *a priori* given coordinate structure,[11] we can now distinguish 'natural' from 'unnatural' properties. This is exactly what Gärdenfors (2000) proposes. In terms of the framework of meaning spaces, called 'conceptual spaces' by him, he called those subsets of an Euclidean meaning space 'natural' properties which form *convex regions* of this meaning space. For a set of objects to be a convex region, or set, it has to be closed in the following sense: if x and y are elements of the set, all objects 'between' x and y must also be members of this set. More formally, if we assume that x and y are points in the meaning space that can be characterized by a vector, then we say that z is a *convex combination* of x and y iff $z = \alpha x + (1 - \alpha)y$, with $\alpha \in [0, 1]$, and where $\alpha\langle x_1, ..., n\rangle = \langle \alpha x_1, ..., \alpha x_n\rangle$. Gemetrically, this means that z is located somewhere in between x and y. It is easy to see that the set of all convex combinations of x and y is the straight line segment joining x and y. A set of vectors C is a *convex set* iff whenever $x \in C$ and $y \in C$, it follows that also any convex combination of x and y is in C, i.e., $\alpha x + (1 - \alpha)y \in C$. Notice that among all the subsets of the meaning space, the set of those that build convex regions of it forms only a very small minority. In consequence, Gärdenfors' proposal that many, if not all, properties denoted by simple natural language expressions form such convex regions of the conceptual space is, potentially, a very strong one. It is appealing as well, because if we know that a set is convex, the extension of the set is much easier to learn than without this knowledge. For instance, in the vector space $\mathbf{R}^2$, the convex set with vertices $\langle 0, 0\rangle$, $\langle 1, 0\rangle$ and $\langle 0, 1\rangle$ contains infinitely many points, but to learn which elements are in this set if one assumes (as a learning bias) that sets are convex, one only has to learn the three edges. Of course, the strength of Gärdenfors' hypothesis crucially depends on what could be the coordinates. For some categories of property denoting expressions, like colors, it is quite clear what the coordinates could be, and Gärdenfors' proposal is quite successful. For other expressions, however, it is less clear whether some-

[11] In denotational semantics, this 'a priori' knowledge can be represented in terms of a restriction on accessible worlds.

thing like the correct coordinate system can be found, and, in consequence, his proposal is much harder to test here. But even if the hypothesis that properties expressed by simple natural language expressions form convex regions of the meaning space is correct, two questions still have to be addressed: (i) in what sense are meaning spaces given *a priori*, and (ii) what makes convex regions so 'natural'? The first question is important, because whether a set of individuals forms a convex region or not crucially depends on the meaning space: in a different meaning space, the same set of objects might not form such a convex region anymore. Perhaps the a priori character of a coordinate system, together with its notion of 'similarity' is just due to a social convention, but perhaps it is, in fact, an innate property of our brain. As even an empiricist like Quine (1968) acknowledges, the latter seems natural for at least some cases, and allows for a standard Darwinian explanations.[12] As for the second question, it is shown in Jäger & van Rooij (2005) that all evolutionarily stable sender-receiver strategy combinations assign descriptive meanings to messages that partition the Euclidean meaning space into convex regions each with a particular point in its center such that all points in a region are closer to its own central point than to any other. Partitions like this are also known as *Voronoi Tesselations*. Thus, Gärdenfors' proposal can be given an evolutionary explanation.

Let me illustrate the evolution of Vorronoi Tesselations by a very simple example. Let us take as our conceptual-, or vector-, space the line segment $[0, 1]$. Each element of $[0, 1]$ is a possible situation from which the sender can send a message. We asume, as before, that the speaker strategy is a function from $T = [0, 1]$ to M, and we assume that M consists only of two messages, m_i and m_j. This means, obviously, that we have many more situations than messages, and thus that in equilibrium there will be at least one message that will be sent in (many) different situations. Assume that P is a probability measure over T which assigns to each point an equal probability. As usual, we take the hearer strategy to be a function from M to $\mathcal{A}$, but because we assume again the existence of a 1-1 correspondence between $\mathcal{A}$ and T, modeled by function f, we might think of the hearer's strategy also as a function from M to T. The goal of the hearer is to 'guess' what is the situation the speaker is in. To account for this goal, we assume that the utility function is defined in terms

[12] 'A standard of similarity is in some sense innate. [...] why does our innate subjective spacing of qualities accord so well with the functionally relevant groupings in nature as to make our inductions tend to come out right? [...] There is some encouragement in Darwin. If people's innate spacing of qualities is a gene-linked trait, then the spacing that has made for the most successful inductions will have tended to predominate through natural selection.' (Quine, 1968, pp 123-126). Recent work of Kirby (2005) might suggest, though, that cultural evolution is involved as well.

of a *similarity measure*, or *distance function*, between points. We will assume that the utility $U(t_i, t_j)$ is higher if the distance between t_i and t_j is lower. Now one can easily see that according to the evolutionarily stable strategies, the *descriptive meanings* of the messages gives rise to the following partition of the state space: the descriptive meaning of the one message, $S^{-1}(m_i)$, will be the *first half* of the line segment, while the descriptive meaning of the other message, $S^{-1}(m_j)$, will be the *second half* of the line segment. Notice that $\{S^{-1}(m_i), S^{-1}(m_j)\}$ is not just a partition of $[0, 1]$, but a very special one in the sense that both cells form convex regions.

Until now we have only looked at descriptive meanings, but what about *imperative meanings*? Recall from section 2 that the imperative meaning of the message sent in t according to language $\langle S, R\rangle$ is $R(S(t))$. On our assumption that $\mathcal{A} = T$, and thus that the receiver strategy is a function from M to T, the imperative meaning will always be a particular point in $[0, 1]$. Which points will that be for our messages m_i and m_j? On our assumption that each point in $[0, 1]$ is equally likely, it will be that $R(m_i) = 0.25$ while $R(m_j) = 0.75$. Notice that $R(m_i)$ is just right in the middle of $S^{-1}(m_i)$, and the same is true for m_j. In fact, we can think of the imperative meanings of the messages as just the *stereotypes* of their descriptive meanings. What's more, once we assume a particular meaning space together with an a priori given distance function (or more in general, a particular utility function), we can *derive* the descriptive meaning of a particular message from its imperative meaning.[13] Indeed, as shown by Gärdenfors (2000), that is one of the beauties of Vorronoi Tesselations. The distinction between descriptive and imperative meaning will play a major role in the following sections as well.

3.2 Compositionality and the meaning of concatenation

Remember that each (evolutionary) equilibrium $\langle S, R\rangle$ was always a solution to a particular coordination problem. For instance, the problem of how to classify a number of objects, or individuals, with respect to color. Of course, there might be another coordination problem involving the same set of objects, but now the problem is how to classify them, e.g., with respect to shape. The combination $\langle S', R'\rangle$ might evolve as the equilibrium for this problem, and also this equilibrium will partition the set T based on a (new, and disjoint to M) set of messages M'. It is now not unreasonable to assume that conjunction is one of the things that might arise the moment we consider *conjoined* coordination problems, in our case of how to classify the objects with respect to color *and* shape.

[13] Of course, the other way round is possible as well.

The natural suggestion is that if the descriptive meaning of $m \in M$ is $S^{-1}(m)$ and of $m' \in M'$ it is $S'^{-1}(m')$, then the combination of the two signals, $m^{\cap}m'$, will denote those objects that have both the property denoted by m *and* the property denoted by m', i.e. $S^{-1}(m) \cap S'^{-1}(m')$.

Notice that the above analysis of conjunctive messages already presupposes that compositional languages can and will emerge. In the previous paragraphs we have explained how the messages of a language can receive a meaning, but those messages were just unstructured wholes. Now we have assumed that messages can be structured, and have suggested how the meanings of these structured wholes can be determined from the meanings of its parts in a compositional way. Why is compositionality so important? A traditional answer is that in this way one can explain why a competent language user is capable of *interpreting* a theoretically infinite number of sentences in finite means.[14] A more interesting answer, perhaps, is that if a language is (assumed to be) compositional, it helps a lot to learn the syntax and semantics of this language: 'any information about either syntax or semantics would provide some evidence about the other, and an optimal learning mechanism would presumably exploit all available evidence' (Partee, 1984, p. 285).

But why and how can such a compositional language emerge in the first place? In the literature there exist two different types of approaches to address this issue. On the first approach, dubbed *synthetic* by Hurford (2000), one assumes that the set of messages is already partitioned – into nouns and verbs, for instance –, with already established meanings, and that new messages emerge from combining two messages from different types, e.g., sentences consisting of nouns and verbs. This approach is adopted by Nowak & Krakauer (1999), and it is shown by using evolutionary game theory that if compositional languages are taken to be more robust against noise than holistic languages, only compositional languages are evolutionarily stable, and thus will emerge. Notice that our above 'story' of conjunction is really in line with the first approach, because it was assumed that the sets of messages M and M' were disjoint, and that all these messages have already separate meanings.

Unfortunately, Nowak & Krakauer's (1999) explanation of the emergence of compositional languages has been criticized on several points, and I believe these points are well-taken. Zuidema (2004), for instance, rightly criticizes the implicitly adopted assumption that we just compare a holistic versus a (pre-existing) compositional language to see which one comes out best (in terms of invasion barrier). By adopting this assumption one already assumes that compositional languages are possible, but does not explain how they can

[14]It should be noted, though, that compositionality is only one way to achieve this goal, which means that the fact that we are successful in interpreting an infinite number of natural language sentences doesn't prove that natural languages are, in fact, compositional.

emerge 'out of' holistic ones. More in general, the *synthetic* point of departure has been claimed to be wrong by Wray (1998), for instance. She claims that the meaningful messages in a holistic language should not be thought of as words, but rather as whole *utterances* that describe not objects of a particular type, but rather particular kinds of *situations*. In the footsteps of the Quinean "radical translation" tradition, she proposes an *analytic* analysis to explain the emergence of compositional languages. According to this kind of analysis, compositional languages do not arise through the combination of meaningful words, but rather through the correlation between (i) features of meaningful messages, and (ii) aspects of the situations that these utterances describe.[15] Notice that if we assume that meanings are represented in an n-ary vector space, we can already take apart a situation, or objects, in various 'aspects'. So let us see how we could work out the analytic approach.

Let us look at a very simple example. Suppose that we have a meaning space consisting of four meanings, given by the vectors $\langle 0,0\rangle$, $\langle 0,1\rangle$, $\langle 1,0\rangle$, and $\langle 1,1\rangle$, and suppose we have a holistic language where, for some reason, all messages are expressions of length 2 using as its first part an element of $\{a,b\}$, and as its second part an element of $\{c,d\}$. It is now clear that, among others, the following two languages, or message-meaning combinations, are possible:

meaning	L_s	L_c
$\langle 0,0\rangle$	ac	ac
$\langle 0,1\rangle$	ad	ad
$\langle 1,0\rangle$	bc	bd
$\langle 1,1\rangle$	bd	bc

We assume here that both L_s and L_c are holistic languages, and because both are equally expressive, for communication they are equally good. If we now adopt the analytic approach towards the emergence of compositional languages, we must be looking for correlations between features of meaningful messages and features of the meanings, such that also the parts of the messages receive an independent meaning. Such a correlation is found easily in the *simple* language L_s: 'a' and 'b' can be thought of as meaning $\{\langle 0,0\rangle,\langle 0,1\rangle\} = \langle 0,i\rangle$ and $\{\langle 1,1\rangle,\langle 1,1\rangle\} = \langle 1,i\rangle$, respectively, while '$c$' always means $\{\langle 0,0\rangle,\langle 1,0\rangle\} = \langle j,0\rangle$, and '$d$' always means $\{\langle 0,0\rangle,\langle 1,1\rangle\} = \langle j,1\rangle$. In the more *complex* L_c, on the other hand, it seems that this kind of correlation cannot be found: although 'a' and 'b' can be thought of as having the same meanings as in L_s, the meanings of 'c' and 'd' seem to depend on whether 'a'

[15] This analytic approach is not limited to the field of language evolution, but can be found also in philosophy of language (e.g. Quine and Davidson) and language acquisition (e.g. Tomassello).

or 'b' was used first. Thus, it seems that L_s is fully compositional, while L_c is not. Although this is not the case in this particular example, compositional languages are in general preferred to holistic languages because they are more robust under a learning bottleneck (e.g. Kirby, 2000) and noisy communicative situations (Nowak & Krakauer, 1999). If we would slightly complicate our example from going to a two- to a three-dimensional meaning space, learning a fully compositional language would only require learning the meaning of 6 symbols, while learning a (completely) holistic language involves learning 9 (or 8) independent message-meaning combinations. Thus, a language like L_s seems to be preferred to a language like L_c because by being compositional it behaves better under a learning-bottleneck. But, then, complicating the semantics of L_c slightly turns also this language in a compositional one: also 'c' and 'd' can be given *context-independent* meanings, if we thinks of meanings in a more abstract way. The meaning of 'c' in L_c, for instance, can be thought of as being a *function* that says: $\langle i, 0\rangle$ if i is 0, and $\langle i, 1\rangle$ if i is 1 (in set terms, this would be $[[c]]_{L_c} = \{\langle\langle 0, i\rangle, \langle 0, 0\rangle\rangle, \langle\langle 1, i\rangle, \langle 1, 1\rangle\rangle\}$, and $[[d]]_{L_c} = \{\langle\langle 0, i\rangle, \langle 0, 1\rangle\rangle, \langle\langle 1, i\rangle, \langle 1, 0\rangle\rangle\}$.) Thus, if we assume that meanings can be slightly more abstract than we thought of before, we can still claim that L_c is a perfectly compositional language. This is only a very simple example, but it seems that languages can usually be tricked this way. Rather than arguing for this by giving more examples, or formal proofs, I will just rely on authority:

> If the syntax is sufficiently unconstrained and meanings are sufficiently rich, there seems no doubt that natural languages can be described compositionally. (Partee, 1984, p. 281)

So, it is not compositionality all by itself that makes languages behave better under a learning-bottleneck. If a language can be analyzed compositionally only by assuming unconstrained syntactic permutation or transformation rules, or by assuming that the meaning of the signs that complex expressions are made of are difficult to learn, or to compute, it doesn't have an evolutionary advantage. Thus, (ignoring syntax) if we take the analytic approach towards the emergence of compositional languages seriously, what computational and learning limitations select for is not so much compositionality, but rather languages that can be given a compositional analysis such that the meanings of the simple expressions are easy to compute, use, or learn.[16]

[16]This point is mostly ignored in much of the work on the evolution of compositional languages. Still, it doesn't make that work nonsensical. In most of this work strong assumptions are (implicitly) made concerning the complexity of syntax and semantics, and if such assumptions are made, whether a language can be analyzed compositionally or not becomes an empirical issue again.

I know of three ways in which complexity can be made to have a selective effect in evolutionary game theory. Either one takes complexity directly into account when defining the utility function, or one assumes that complexity functions as a filter for successful communication, which only indirectly influences the utility of a language, or one assumes that complexity influences learning. Let me just give a *very* simplifying but high-level analysis here. Assume that we restrict ourselves in our language game to only two languages, L_1 and L_2. If these languages are used by different agents, this might give rise to the following payoff table:

$\mathcal{U}(L_i, L_j)$	L_1	L_2
L_1	1,1	$0, 0$
L_2	$0, 0$	1,1

What this table models is not only that users of different languages don't understand each other, but also that as far as communication is concerned, it doesn't matter which language is used. On these assumptions it follows in the replicator dynamics behind evolutionary game theory that language L_1 grows if and only if the majority of the population (however tiny this majority is) uses L_1. What if we give up our two assumptions? Then we end up with the following table:

$\mathcal{U}(L_i, L_j)$	L_1	L_2
L_1	a, a	c, d
L_2	d, c	b, b

We have seen in section 2 that in the replicator dynamics behind evolutionary game theory L_1 grows whenever the distribution of players that plays L_1 is higher than $\frac{(c-b)}{-(a+b)+(c+d)}$. If we now still assume that users of different languages don't understand each other, it follows that $c = d = 0$, and language L_1 grows whenever the distribution of players that plays L_1 is higher than $b/(a + b)$. Thus, if a is higher than b, $a > b$, L_1 has a better chance to grow in a population of players than L_2. Technically it means that L_1 has a greater 'basin of attraction'. But why should b be lower than a? One reason might be that using L_2 to communicate about the world is computationally more complex, and that this complexity has an immediate payoff-relevant penalty (e.g. van Rooij (2004a,b), Jäger (ms)). Another reason might be that using L_2 is more complex and thus less successful under noisy situations (e.g. Nowak & Krakauer, 1999). If one assumes a noisy communication channel, $U(t, L_i, L_j)$ should not simply be defined as $\frac{1}{2}U(t, R_j(S_i(t))) + \frac{1}{2}U(t, R_i(S_j(t)))$. Rather, the message send by S_i (or S_j) in t can be distorted, and thus that R_j (or R_i) applies to the distortion of $S_i(t)$ (or $S_j(t)$), modeled by a distortion ma-

trix. Another reason why L_1 might be 'preferred' (in the sense of being more likely to emerge and harder to invade) to L_2, of course, is that L_2 is harder to learn. This, too, can be represented in evolutionary game theory, at least if we slightly complicate the dynamics 'behind' the evolutionary stable states (cf. Komarova et al (2001)). In the standard replicator dynamics, the proportion of players that uses language L_i in a particular generation depends only on the proportion of players playing L_i in the previous generation, and the difference between (i) the average utility of playing L_i in this previous generation and (ii) the average utility of the whole population. Once we take learning into account, we don't have to assume anymore that all 'children' of agents playing L_i also play L_i: if L_i is difficult to learn, not all, or even just a few 'children' of agents using L_i will perfectly acquire this language. This will have the result that L_i won't have a very good chance to become the shared language of the members of a population.

We have seen above that not only L_s but also L_c can be analyzed such that all the expressions have context-independent meanings. However, there was still a difference between the independent meanings of 'c' in the two languages: the meaning of 'c' in L_s is the same as its denotation, but this is not the case for the meaning of 'c' in L_c. In L_c, the denotation of 'c' depends not only on its (context-independent) meaning, but also on the context in which it is used (whether it is used in the context of 'a' or in the context of 'b'). Thus, what evolution seems to select for is not languages consisting of expressions with a context-independent meaning (for that can almost always be arranged), but rather for languages consisting of expressions whose *denotations* are *context-independent*, i.e., whose denotations equal their meanings.[17]

Think of the languages L_s and L_c as subject predicate sentences, or, for instance, as adjective noun combinations. Taking the first alternative, we can think of $\{a, b\}$ as subject expressions and of $\{c, d\}$ as predicate expressions. In each language we know already what the meanings are of all potential subject predicate combinations, but how can they be determined in terms of the meanings of their parts? In L_s things are very simple: the meaning of a sentence of the form $x^\frown y$, $[[x^\frown y]]_{L_1}$, is just $[[x]]_{L_s} \cap [[y]]_{L_s}$. In L_c, however, we have to make use of a more general mode of semantic combination: *functional application*: $[[x^\frown y]]_{L_c} = [[x]]_{L_c} \cap [[y]]_{L_c}([[x]]_{L_c}) = [[y]]_{L_c}([[x]]_{L_c})$. Thus, L_s seems to be preferred to L_c because in the semantics for the latter we require this more general mode of semantic combination, while in the former we don't. This point is not different from what we noted above, although it perhaps highlights

[17] This cannot be the whole story, of course, because for reasons of efficiency, all languages make use of personal pronouns and tenses, for instance, whose denotations are very context-dependent. Still, it remains true that expressions are selected for whose denotations depend on context in very *predictable* ways.

once more how costly some standard analyses in denotational semantics really are.

An argument for descriptive meaning? Assume, as before, that all messages are expressions of length 2 using as its first part an element of $\{a, b\}$, and as its second part an element of $\{c, d\}$. But let us now look at a completely different meaning space: the meanings are points in the line segment $[0, 1]$. What will be the meanings of the 4 complex expressions? It is easy to see that this will depend on the probability distribution over $[0, 1]$. Let us first consider the case where all points are equally likely. In that case, the following type of meaning assignment is natural:

L_1	imperative meaning	descriptive meaning
ac	0.125	$[0, 0.25]$
ad	0.375	$[0.25, 0.5]$
bc	0.625	$[0.5, 0.75]$
bd	0.875	$[0.75, 1]$

After analyzing this language, we can also give imperative and descriptive meanings to the 4 simple symbols, such that concatenation is analyzed as imperatively meaning '+', and as descriptively meaning '$\cap$'. There are several ways to do this[18], but if we assume that L_1 emerged from a simpler language that described the line segment using only a and b as messages, only the one below will result:

L_1	imperative meaning	descriptive meaning
a	0	$[0, 0.5)$
b	0.5	$[0.5, 1]$
c	0.125	$[0, 0.25) \cup [0.5, 0.75)$
d	0.375	$[0.25, 0.5) \cup [0.75, 1]$

Now assume that the probability is not equally distributed over $[0, 1]$, but that it gives rise to, for instance, a normal distribution with a peak at 0.5 and minima at 0 and 1. Then, something like the following meaning assignment will follow:

L_2	imperative meaning	descriptive meaning
ac	0.25	$[0, 0.35)$
ad	0.4	$[0.35, 0.5)$
bc	0.6	$[0.5, 0.65)$
bd	0.75	$[0.65, 1]$

[18]Thanks to an anonymous reviewer on this point.

After analyzing this language, we can also give imperative and descriptive meanings to the 4 simple symbols, such that concatenation is analyzed as imperatively meaning '+', and as descriptively meaning '$\cap$':[19]

L_2	imperative meaning	descriptive meaning
a	0	$[0, 0.5)$
b	0.5	$[0.5, 1]$
c	0.25, if a, 0.1, if b	$[0, 0.35) \cup [0.5, 0.65)$
d	0.4, if a, 0.25, if b	$[0.65, 0.5) \cup [0.65, 1]$

What this simple example might suggest is that meanings should preferably be thought of descriptively, rather than imperatively, and thus why concatenation should mean '$\cap$' rather then '+'. The reason is that although the two come down to the same if the meanings, or points in $[0, 1]$, are equally distributed, the descriptive meaning-analysis combined with a Boolean analysis of concatenation seems to allow for simpler meanings of the parts if the probability distribution is not uniform, and the language is still analyzed compositionally. Our example involved only a very simple meaning space, but it is easy to see that the same would follow if we consider more complicated meaning spaces (such as the binary meaning space $[0, 1] \times [0, 1]$), or when the meaning space does not give rise to ordered coordinates in the first place. Thus, a Boolean analysis of concatenation seems to be preferred, because in general it allows for simpler, or context-independent, meanings of its parts, when the language is interpreted compositionally.

Perhaps unfortunately, however, this argument is not as convincing as it might seem. It is not really clear why the descriptive meanings of 'c' and 'd' are simpler than their imperative meanings. Notice first that if the best way to represent a meaning is in terms of a set, it means that its meaning is just an *enumeration* of instances. So, what we are after is compact representations of this set. Although at first sight the descriptive meanings of symbol 'c' in L_1 and L_2 are equally complex – just the union of two sets of points –, in L_1 the descriptive meaning of 'c' can be represented more compactly than in L_2 in terms of how to *compute* this descriptive meaning: it means just something like 'the first half of'. Something like this cannot be done as easily for the meaning of 'c' in L_2. So, also the descriptive meaning of 'c' in L_2 is perhaps not as simple as it seemed at first, and it is not clear why this meaning is any simpler than the imperative meaning of 'c' in L_2. Thus, our suggested motivation for the primacy of descriptive meaning, combined with a Boolean

[19] Assuming again that L_2 emerged from a language that used only $\{a, b\}$.

analysis of concatenation, is not as convincing as it might have seemed.[20]

Still, once we want to determine the meaning of a complex expression from the meanings of its parts, it is generally assumed that this cannot be done by looking only at imperative meanings, or stereotypes. What should the mode of combination be if stereotypes are modeled by vectors? Addition, or convex combination will not do in most cases. Still, Gärdenfors (2000) suggests that there might be another mode of combination after all. Consider adjectives-noun combinations like *stone lion*. It is clear that analyzing the meaning of this complex phrase by means of intersection of the descriptive meanings of its parts won't give the correct result. A standard conclusion out of this in denotational semantics is to assume that the adjective and the noun should not be given meanings of the same type, and that concatenation should be interpreted as functional application. But, as argued for by Gärdenfors (2000), we don't always need such a computational complex analysis which requires the use of higher order logic. Assume that the adjective and the noun have meanings of a different type in the sense that whereas the *imperative* meaning of the noun is a vector with two coordinates, characterized in terms of numbers on both the x-axis (say, material) and the y-axis (e.g. shape), the imperative meaning of the adjective is a vector with only one coordinate, e.g., only characterized in terms of a number on the x-axis. Gärdenfors proposes now *substitution* as a general mode of interpretation for concatenation: the (imperative) meaning of 'stone lion' is the same as the imperative meaning of 'lion', but with the number of the x-axis replaced by the imperative meaning of 'stone'. Thus, 'stone lion' denotes an object with the stereotypical shape of a lion, but that is made of stone. It is clear that substitution can account for some meanings of concatenation where functional application has been used in denotational semantics. It remains unclear, however, how general it really is.

3.3 Truth conditional connectives

Disjunction Under which circumstances can a language evolve in which we have a message that means 't_i', one that means 't_j', and yet another with the disjunctive meaning 't_i or t_j'? We have seen above that if there exists a 1-1 function, f, from situations to (optimal) actions to be performed in those

[20]Notice that our discussion also suggests that if we want to think of meanings from an evolutionary perspective, we shouldn't think of them in the first place as sets, as in denotational semantics. Rather, we should think of them in *richer* ways such that we can determine how difficult it is to compute or learn a given meaning. Notice that this points in the direction of *procedural semantics*, but it also suggests that the imperative, or stereotypical, meaning of a simple expression is more basic than its denotational meaning.

situations, a language can evolve with a 1-1 correspondence between messages and meanings. The existence of this function f, however, won't be enough to 'explain' the emergence of messages with a disjunctive meaning. What is required, instead, is a 1-1 function from *sets* of situations to (optimal) actions. We can understand such a function in terms of a payoff table like the following:

$U(t,a)$	a_1	a_2	a_3	a_4	a_5	a_6	a_7
t_1	4	0	0	3	3	0	2.3
t_2	0	4	0	3	0	3	2.3
t_3	0	0	4	0	3	3	2.3

Notice that according to this payoff table, action a_1 is the unique optimal action to be performed in situation t_1, and the same holds for combinations $\langle t_2, a_2\rangle$ and $\langle t_3, a_3\rangle$. So far, this is the same as what we had in section 2. This table, however, contains more information. Suppose that the speaker (and/or hearer) knows that the actual situation is either t_1 or t_2, and that both situations are equally likely. In that case the best action to perform is neither a_1 nor a_2 – they only have an expected utility of 2 –, but rather a_4, because this action now has the highest expected utility, i.e., 3. Something similar holds for information 't_1 or t_3' and action a_5, and for 't_2 or t_3' and action a_6. Finally, in case of no information, which corresponds with information 't_1 or t_2 or t_3', the unique optimal action to perform is a_7. Thus for all (non-empty) subsets of $\{t_1, t_2, t_3\}$ there exists now a unique best action to be performed. Notice that each such subset may be thought of as an *information state*, the (complete or incomplete) information an agent might have about the actual situation. Suppose now that we lift the sender-strategy from a function that assigns to each *situation* a unique message to be sent, to one that assigns to each *information state* a unique message to be sent. It is not difficult to see that now we will end up (after evolution) with a communication system in which there exists a 1-1-1 correspondence between information states (or sets of situations), messages, and actions to be performed (cf. Skyrms, 2004).[21] Thus, there will now be messages which have a disjunctive meaning. This by itself doesn't mean yet that we have a separate message that denotes disjunction, but only that we have separate messages with disjunctive meanings in addition to messages with simple meanings. However, as convincingly shown by Kirby (2000) and others, under a learning bottleneck, languages are forced to become compositional. Given that in all evolutionarily stable linguistic conventions there will be separate messages that denote the most informative information states – i.e. for all situations t_i there will always be a message with (descriptive)

[21] This is a general result if there is a 1-1 correspondence between sets of situations and actions, and not restricted to the particular example discussed above.

meaning $\{t_i\}$ –, what will evolve under iterated learning forced by a learning bottleneck is a complex message with meaning $\{t_i, t_j\}$ that consists of three separate signals: one signal denoting $\{t_i\}$, one signal denoting $\{t_j\}$, and one signal that turns these two meanings into the complex meaning $\{t_i, t_j\}$, which is done by (set theoretical) *union*. The latter signal, of course, might then be called 'disjunction'.

Although there appears to be some contrasting evidence, it is standardly assumed that humans, and only humans, have (compositional) languages containing messages which have a disjunctive meanings. In terms of the above result we might speculate why this is the case. Recall that for disjunctive messages to evolve, it was crucial that we took *information states*, or *belief states* into account. We might now speculate that it is the existence of such belief states that sets us apart from (other) animals, and why we, but not them, could make use of messages with disjunctive meanings.

In principle, once we take information states into account, we can not only state under which circumstances disjunctive messages will evolve, but also when negative and conjunctive messages will evolve. The main difference is that we have to assume more structure of the set of information states (for disjunction we only required that this set is an i-join semi-lattice, for both disjunction and conjunction to evolve we most naturally have to assume that the set forms a full lattice, while for negation to evolve as well we even have to assume that the set is a free lattice, which contains much more elements, and is of a more complicated structure than a lattice, and certainly than a i-join semi-lattice).

Notice that there is a distinction between the circumstances under which disjunction, and under which conjunction can arise. For disjunction to arise, we needed the existence of information states: we required that the sender has, and knows that he has, incomplete information about the actual situation, and we had to 'lift' the speaker strategy to a function from information states to messages. In that case, messages will evolve that have a disjunctive meaning and denote *sets* of situations. For conjunction to make sense, we also had to assume that there will be messages that (descriptively) denote sets of situations. However, for that to be the case we didn't require that a speaker strategy is a function taking sets of situations as input: it can just be a function that maps situations to messages. In that sense, conjunction is 'simpler' than disjunction.[22] There is another sense in which conjunction is 'simpler' than (standard) disjunction, and this involves the notion of 'convexity'. This might have evolutionary impact as well.

[22]In fact, in natural language it seems that we don't really need conjunctions to conjoin messages: sequencing is good enough. According to Gil (1991) there are indeed (modern) languages that don't have the notion of a truth-conditional conjunction at all.

Above I have assumed that disjunction has its standard Boolean meaning and only investigated under which circumstances can expressions evolve with this meaning. But why should we make this assumption, and why shouldn't we take into account the historical roots of expressions which now have the standard meaning? As for the assumption, perhaps because meanings correspond with 'innate' Boolean laws of thought. But, then, even if disjunction has an innate Boolean meaning, we would like to explain how this could have evolved.

It is widely assumed that in the animal kingdom what counts is not descriptive meaning, but rather imperative meaning. Moreover, this imperative meaning is just an action. In manipulative communicative situations, however, it might be good to leave the addressee a *choice* between several alternatives, and it might be that 'or' is used first to signal this. This would mean that if the imperative meanings of X and Y are x and y respectively, the imperative meaning of 'X or Y' could be described as $\alpha x + (1 - \alpha)y$, with $\alpha \in \{0, 1\}$. Now suppose that later in evolution, the natural meanings of X and Y are descriptive, rather than imperative, and that both denote *sets* of objects/situations. In that case, the meaning of 'X or Y' could be the following set $\{\alpha x + (1 - \alpha)y : x \in X, y \in Y, \alpha \in \{0, 1\}\} = X \cup Y$.

Perhaps the last step was too quick. Recall that *convexity* is both a useful constraint on properties, and one whose emergence can be explained. What this suggests is that as many as possible properties that we use should denote convex sets. Unfortunately, if we assume that disjunction is interpreted as set theoretical union, in contrast to intersection, the union of two convex sets need itself not be convex. So we would like to explain the emergence of our non-convex notion of disjunction from a closely corresponding convex notion of disjunction. If we assume that the meaning space has the form of a vector space, I think we can give such an appealing explanation. The reason is that there is another natural operation that can be associated with disjunction that results in convex sets. This operation is just that of *convex combination*. Remember that a convex combination of two vectors x and y is any vector $\alpha x+(1-\alpha)y$ with $\alpha \in [0, 1]$. This operation can be lifted naturally from vectors to *sets* of vectors in the following way: if X and Y are sets of vectors, we define the convex combination of X and Y, $X + Y$, as $\{\alpha x + (1 - \alpha)y | x \in X, y \in Y, \text{ and } \alpha \in [0, 1]\}$. In distinction with Boolean disjunction, $X+Y$ might have elements that are neither in X nor in Y.[23] One can show that if X and Y both denote convex sets, what $X + Y$ gives us is the smallest convex set that

[23]See Widdows & Peters (2003) for some uses of 'or' that perhaps can be analyzed in terms of $X+Y$. Widdows (2004) argues that what is denoted by 'mammals' can be thought of as the convex closure of the denotations of 'rabbits', 'pigs' and 'dolphins', and contains many creatures that do not belong to any of the three sets.

contains both X and Y. If convexity is a strong constraint not only on 'simple' properties, but on 'complex' ones as well, I believe that convex combination is a natural candidate for being a meaning of disjunction. It should be clear that the standard Boolean meaning of 'X or Y' is just the same as $X + Y$, except that α should be an element of $\{0, 1\}$ instead of an element of $[0, 1]$. Can we assume that the former evolves out of the latter? I think we can. Notice that 'convex combination' is, in general, an analog, or continuous, notion. It is a relatively standard assumption in epistomology (e.g. Dretske, 1981) that the crucial difference between perceptual and cognitive phenomena is that the first, but not the latter, is mostly analog in character: cognitive information is normally *discrete* or *digital*. According to Sebeok (1962), a similar difference exists between animal and human communication systems. What I would like to suggest is that the step from interpreting disjunction as convex combination to interpreting it in the standard way is just this step from analog to discrete. In our case, it would mean that α should be an element of $\{0, 1\}$ instead of an element of $[0, 1]$. As we have seen above, what results is $X \cup Y$, i.e., set-theoretical, or Boolean, union.

Why not more connectives? Above we have given some motivation for why, and under which circumstances, certain truth-functional connectives could arrive. We have motivated the existence of one unary truth-functional connective, negation, and two binary truth-functional connectives, disjunction and conjunction. However, once we assume that each (declarative) sentence is either true or false, there are *four* potential unary connectives, and as much as *sixteen* potential binary connectives. Although all these potential connectives *can* be expressed in natural language, the question is why only one unary and only two (or perhaps three) binary truth-functional connectives are expressed by means of simple words in all (or most) natural languages? That is, can we give natural reasons for why languages don't have the truth-functional connectives that are mathematically possible? Fortunately for us, this problem has already been solved convincingly by Gazdar & Pullum (1976). To illustrate their proposal, let us look at the four possible unary connectives, $c_1, ..., c_4$:

p	$c_1\ p$	$c_2\ p$	$c_3\ p$	$c_4\ p$
1	0	1	0	1
0	1	0	0	1

Connective c_1 is, of course, standard negation. Why don't we see the others in natural language(s)? The second one obviously doesn't make a lot of sense: $c_2\ p$ just has the same truth-value as p itself, is thus superfluous, and can be explained not to exist in this way. Connectives c_3 and c_4 are equally strange: the truth values of $c_3\ p$ and $c_4\ p$ are independent of the truth value of p.

Assuming, as a *strict compositionality* requirement, that all arguments of a connective have to be potentially relevant to determine the truth value of the whole, c_3 and c_4 are ruled out. Thus, only c_1 is left, just as desired.

Gazdar & Pullum (1976) show that when we (i) assume the above principle of *strict compositionality*, (ii) require (binary) connectives to be *commutative*, and (iii) assume a principle of *confessionality*, which forbids natural languages to have any (binary) connective which yields the value true when all its arguments are false, then all potential binary connectives are ruled out except for the following three: conjunction, standard (inclusive) disjunction, and what is known as *exclusive* disjunction. This is an appealing result, given that strict compositionality makes perfect sense (a language that doesn't obey it is not efficient), the principle of confessionality can be explained by the difficulty of processing negation, while the constraint of commutativity is motivated by the not unnatural idea that the underlying structures of the connected sentences are linearly unordered. But it still leaves us with exclusive disjunction, a connective of which many people have argued that it is not expressed by a simple word in any language.[24] Fortunately, there are several ways to rule out this connective as well: based on the observation that when taking more than 2 arguments, exclusive disjunction gives rise to unnatural predictions ('p or q or r'‘ is predicted to be true when either exactly one of the arguments is true, or all of them!), Gazdar & Pullum show that it is ruled out by generalizing their analysis by assuming that connectives take sets rather than sequences of truth values as arguments. Horn (1989), on the other hand, argues that the existence of a connective expressing exclusive disjunction is not required anyway, because this meaning already follows from standard inclusive disjunction in combination with the scalar implicature that not both disjuncts are true.

3.4 Relations and prepositions

Just as for connectives and properties, also for relations it is the case that many more relation-meanings *can* be expressed between a number of objects, than that we typically express by simple natural language expressions. And because there are many more subsets of $D \times D$ (the set of denotations of relational expressions) than that there are subsets of D (the set of denotations of property expressions), or subsets of $\{1, 0\} \times \{1, 0\}$ (the set of binary truth-functional connectives), the problem is much more serious here, but also more difficult to solve. But also here utility, learnability and complexity seem to be important factors.

First, utility plays obviously an important role, and this can again, to some extent, be explained in terms of our signaling game perspective. When we think

[24] Although others have argued that Latin *aut* does express exactly this connective.

of the meaning of a relation as a set of ordered pairs, the sender strategy, S, is now a function from ordered pairs of situations, or objects, $\langle t, t' \rangle$, to messages, while the receiver's strategy, R, one from messages to actions. In analogy to what we saw for properties, it holds that if there are either less messages or less actions than there are ordered pairs of situations, at least some messages will denote, in equilibrium, sets of ordered pairs.

Consider the case of four individuals consisting of a man, his wife, and their two children, a boy, and a girl. Given that we have 4 individuals, we have $4 \times 4 = 16$ ordered pairs. What is an example of a natural partition of this set of ordered pairs? A natural partition could be, for instance, the following set of relational-expressions:[25] {*father-of, mother-of, husband-of, wife-of, son-of, daughter-of, brother-of, sister-of, identical-to*}.[26] Another example of a natural partition would be the relation of individuals situated on a length-scale into {*longer than, equally long as, shorter than*}.

The latter example suggests already that Gärdenfors' (2000) analysis of 'natural properties' can sometimes be extended to 'natural relations'. To do so, Gärdenfors proposes that one dimension of the meaning space measures the length of individuals. The comparative relation *longer than* would then be represented by all ordered pairs of points in the space defined by this dimension. How can we think of this relation as denoting a convex set? Well, we might think of a new binary meaning space where the first and second dimension measure the length of the first and second object of each ordered pair, respectively.[27] In that case, the set of ordered pairs that constitutes the denotation of the relation 'longer than' forms a convex region of the above mentioned meaning space. In fact, this holds for all kinds of comparative relations where the generating dimension is isomorphic to the set of all (positive) real numbers (e.g. *larger*, *earlier*, and *before*). For other relation-denoting expressions we can think of a relation R as a function that takes an object x and maps it to the set of objects y such that xRy. Consider, for instance, a *locative preposition* like 'in front of'. Combined with a noun phrase, this preposition denotes the set of objects in front of the denotation of the noun phrase. In Zwart's (1997) vector-based semantics, objects can be thought of as vectors, and so the set of objects in front of the denotation of the noun phrase is a set of vectors as well. Zwarts (1997) states three semantic universals that involve

[25] Certainly if we disregard the last relation, which is there for technical reasons, only.

[26] In an extensive discussion of semantic universals, Leech (1974) suggests that because all basic kinship relations of (all) languages can at least be expressed in traditional componential analyses as that of Lounsbury (1956) and others in terms of the relations in this set, that this set of kinship relations might be thought of as a universal categorization of kinship relations.

[27] Gärdenfors (2000, p. 92) attributes this idea to Holmqvist.

locative prepositions: (i) the set of vectors denoted by any simple locative preposition applied to an object is closed under shorthening;[28] (ii) this set of vectors is both linearly, and (ii) radially *continuous*.[29] What matters here is that all three universals follow immediately if these sets of vectors (or objects) are taken to denote convex subsets of the meaning space.[30] Of course, proposing that also 'natural relations' are those sets of n-tuples that form, from a certain perspective, a convex region of a meaning space is interesting from our perspective, because we have seen that we can give a natural evolutionary motivation for the notion of convexity.

Convexity can be used to constrain meanings of simple relational expressions in other ways as well. In the footsteps of generative semanticists, Dowty (1979) proposes that many verb meanings can be decomposed in terms of meanings of stative predicates plus some abstract notions like CAUSE and BECOME whose meaning is defined in his aspect calculus. The transitive verb *open*, for instance, is decomposed in terms of the stative predicate 'being open' as follows: $\lambda x \lambda y[CAUSE(x, BECOME(be - open(y))$. Dowty suggests that we should exclude predicates whose interpretation depends on the state of the world at more than one time (or in more than one possible world) in any way other than in the ways explicitly allowed for by the tense and modal operators of his calculus. This by itself does not constrain enough the possible verb meanings, but – as explicitly suggested by Dowty (1979, section 2.4) – it would be a strong constraint if we now assume that the stative predicates (or at least the stage-level ones) should only be predicates that denote (convex) regions of logical space.

Recall that the assumption of convexity is of considerable help to determine (learn or compute) the extension of a set, or relation. The reason is that convexity is a very strong closure condition. But relations might be closed under other conditions as well, and this will also help to determine their extensions.[31]

[28] A region of vectors R is closed under shortening iff for every $\mathbf{v} \in R, s\mathbf{v} \in R$, for every $0 < s < 1$, where s is a scalar (Zwarts, 1997).

[29] A vector $\mathbf{v}$ is *linearly between* $\mathbf{u}$ and $\mathbf{w}$ if $\mathbf{v}$ is a lengthening of $\mathbf{u}$ and $\mathbf{w}$ is a lengthening of $\mathbf{v}$. A vector is *radially between* two vectors $\mathbf{u}$ and $\mathbf{w}$ that from an acute angle if the shortest rotation of $\mathbf{u}$ into $\mathbf{w}$ passes over $\mathbf{v}$. A region of vectors is linearly/radially *continuous* iff for all $\mathbf{u}, \mathbf{v} \in R$, if $\mathbf{v}$ is linearly/radially between $\mathbf{u}$ and $\mathbf{w}$, when $\mathbf{v} \in R$ (Zwarts, 1997).

[30] The idea to relate the meaning of locative prepositions with convexity was explicitly mentioned in later work of Zwarts, and also discussed in Gärdenfors (2000).

[31] Notice that if we can already determine the length of an object, it is easy to determine whether one object is longer than another. To determine the meaning of the adjective 'long', however, more seems to be needed: to compute whether an object x is 'long' we have to compare the set of objects that are longer than x with the set of objects that are not longer than x, which involves much more computational recourses than determining the comparative relation. This might be an argument for why adjectives evolved later than the comparative relation, if that is, in fact, the case.

For instance, a relation might have the higher order property of being *reflexive*, *symmetric*, *transitive*, etc. It is obvious that once we know that a relation has certain of these 'natual' higher order properties, it becomes much easier for agents to learn and remember the extension of this relation. If you know, for example, that a relation R is reflexive, you don't need to check any object to know that this object bears the relation R to itself, and if you know that a relation R is symmetric, learning that x stands in relation R to y suffices to know that also y stands in relation R to x. From this point of view one would expect that those relations that are expressed a lot by simple natural language expressions are such that they have many of such natural 'higher order' properties. I don't really know, to be honest, whether this is the case, but I *do* know that some simple relation-denoting expressions that we seem to find in all languages do have many such properties. Many relation-denoting expressions, for instance, denote *symmetric* and *irreflexive* relations, e.g. 'opposite to', 'near to', 'be married to', 'similar to'.[32] Many other expressions denote ordering relations, relations that are *asymmetric*, *irreflexive*, and *transitive*. Examples are 'above' and 'below', 'before' and 'after', 'in(side)', and all comparative relations. A *linear relation* is any ordering relation that has the additional property of being *connected*: for any x and y, either xRy, or yRx (or $x = y$). In a very interesting article, the economist A. Rubinstein (1996) shows that linear orderings are optimal with respect to learning, in the sense that the minimal number of observations is required in order to learn the extension of the relation. Moreover, he shows that linear orderings are optimal in terms of *expressibility*: if you know that a set of objects stands in a particular relation R to each other, the best relation that this could be is a linear relation, because then we can denote any element of the set in terms of R (plus the logical expressions) only.

4 Conclusion

In this paper we suggested some evolutionary motivations for some proposed semantic universals using game theory. Our motivations made use of notions like *utility*, *learnability*, and *complexity*: we expect those meanings to be universally expressed in simple terms that are useful, easy to learn and remember, and easy to use. We suggested some ways in which these notions have evolutionary bite, and how they might have given rise to semantic universals. In the future we would like to see how our work on linguistic universals is connected with work done in Edinburgh (e.g. Kirby, 2000) by modeling biases for learning that have semantic influence.

[32] On the assumption, at least, that one can not be near to, or similar to, oneself.

References

[1] Bickerton, D. (1981), *Roots of Language*, Karoma Publishers.

[2] Donaldson, M., M. Lachmann & C. Bergstrom, 'The evolution of functionally referential communication in a structured world.', manuscript.

[3] Dowty, D. (1979), *Word Meaning and Montague Grammar*, D. Reidel, Dordrecht.

[4] Dretske. F. (1981), *Knowledge and the Flow of Information*, MIT Press, Cambridge, Mass..

[5] Fraassen, F. van (1967), 'Meaning relations among predicates', *Nous*, **1**: 161-179.

[6] Gärdenfors, P. (2000), *Conceptual Spaces. The Geometry of Thought*, MIT Press, Cambridge, MA.

[7] Gazdar, G. & G.K. Pullum (1976), 'Truth-functional connectives in natural language', *Papers from the the 12th Regional Meeting, Chicago Linguistic Society*, pp. 220-234.

[8] Gil, D. (1991), 'Aristotle goes to Arizona, and finds a language without 'and', In. D Zaefferer (ed.), *Semantic Universals and Universal Semantics*, Foris, Dordrecht, pp. 96-130.

[9] Goddard, C. (2001), 'Lexico-semantic universals: A critical overview', *Linguistic Typology*, **5**: 1-65.

[10] Horn, L.R. (1989), *A Natural History of Negation*, University of Chicago Press, Chicago.

[11] Hurford, J. (1989), 'Biological evolution of the saussurian sign as a component of the language acquisition device', *Lingua*, **77**: 187-222.

[12] Hurford, J. (2000), 'The Emergence of Syntax', In: C. Knight, M. Studdert-Kennedy and J. Hurford (eds.), *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*, (editorial introduction to section on syntax), Cambridge University Press. pp. 219-230.

[13] Huttegger, S. (to appear), 'Evolution and the explanation of meaning', to appear in *Philosophy of Science*.

[14] Jäger, G. (ms), 'Evolutionary game theory and linguistic typology: a case study', to appear in *Language*.

[15] Jäger, G. and R. van Rooij (to appear), 'Language structure: psychological and social constraints', *Synthese*.

[16] Kirby, S. (2000), 'Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners'. In: C. Knight (ed.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, Cambridge University Press, pp. 303–323.

[17] Kirby, S. (2005), 'The evolution of meaning-space structure through iterated learning', In A. Cangelosi and C. Nehaniv (eds.), *Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communcation*, pp. 56-63.

[18] Komorova, N.L, P. Niyogi, and M. Nowak (2001), 'The evolutionary dynamics of grammar acquisition', *Journal of Theoretical Biology*, **209**: 43-59.

[19] Leech, G.L. (1974), *Semantics*, Penguin Books, Ltd, Middlesex, England.

[20] Lewis, D. (1969), *Convention: A Philosophical Study*, Harvard University Press, Cambridge, Massachusetts.

[21] Lounsbury, F.G. (1956), 'A semantic analysis of Pawnee kinship usage', *Language*, **32**: 158-1994.

[22] Maynard-Smith, J & G.R. Price (1973), 'The logic of animal conflict', *Nature*, **146**: 15-18.

[23] Maynard-Smith, J. (1982), *Evolution and the Theory of Games*, Cambridge University Press, Cambridge.

[24] Nowak, M. and D. Krakauer (1999), 'The evolution of language', *Proc. Natl. Acad. Sci. USA*, **96**: 8028-8033.

[25] Partee, B. (1984), 'Compositionality', In F. Landman and F. Veltman (eds.), *Varieties of Formal Semantics*, pp. 281-311. Dordrecht; Foris.

[26] Pawlowitsch, C. (ms), 'Why evolution does not always lead to an optimal signaling system', University of Vienna.

[27] Quine, Q.V. (1969), 'Natural kinds', in *Ontological Relativity and Other Essays*, Columbia University Press, New York, pp. 114-138.

[28] Rooij, R. van (2004a), 'Signaling games select Horn strategies', *Linguistics and Philosophy*, **27**: 493-527.

[29] Rooij, R. van (2004b), 'Evolution of conventional meaning and conversational principles', *Synthese*, **139**: 331-366.

[30] Rubinstein, A. (1996), 'Why are certain properties of binary relations relatively more common in natural language?', *Econometrica*, **64**: 343-356.

[31] Sebeok, T. A. (1962), 'Evolution of signalling behavior', *Behavioral Science*, **7**: 430-442.

[32] Skyrms, B. (2001), *Evolution of the Social Contract*, Cambridge University Press, Cambridge, Massachusetts.

[33] Skyrms, B. (2004), *The Stag Hunt and the Evolution of Social Structure*, Cambridge University Press, Massachusetts.

[34] Stalnaker, R. (1981), 'Anti-essentialism', *Midwest Studies in Philosophy*, **4**: 343-355.

[35] Taylor, P. & L. Jonker (1978), 'Evolutionary stable strategies and game dynamics', *Mathematical Biosciences*, **40**: 145-56.

[36] Wärneryd, K. (1993), 'Cheap talk, coordination, and evolutionary stability', *Games and Economic Behavior*, **5**: 532-546.

[37] Wason, P. C. (1959), 'The processing of positive and negative information', *Quarterly Journal of Expermental Psychology*, **11**: 92-107.

[38] Widdows, D. (2004), 'Geometrical ordering of concepts, logical disjunction, and learning by induction', *Compositional Connectionism and Cognitive Science*, AAAI Fall Symposium Series, Washington.

[39] Widdows, D. & S. Peters (2003), 'Word vectors and quantum logic. Experiments with negation and disjunction', in T. Oehrle & J. Rogers (eds.), *Mathematics of Language 8*, Bloomington, Indiana.

[40] Zuidema, W. (2004), *The major transitions in the evolution of language*, PhD thesis, Edinburgh.

[41] Zwarts, J. (1997), 'Vectors as relative positions: A compositional semantics of modified PPs, *Journal of Semantics*, **14**: 57-86.