



# Applications of Game Theory in Linguistics

Gerhard Jaeger\*

University of Bielefeld

---

## Abstract

The article gives a brief overview over the budding field of game theoretic linguistics, by focusing on game theoretic pragmatics on the one hand, and the usage of evolutionary game theory to model cultural language evolution on the other hand. Two specific applications are discussed in detail: the derivation of scalar implicatures via rational reasoning over communicative strategies, and the predictive potential of an evolutionary interpretation of exemplar dynamics in phonetics.

---

## 1 Introduction

Game theory is a branch of applied mathematics that models situations of strategic interaction between several agents. Since its inception in mid-twentieth century (von Neumann and Morgenstern 1944), it has developed into a standard tool in economics, and it is widely used in other academic disciplines as well, especially in political science, biology, and philosophy.

At least since Lewis 1969 and Spence 1973, the strategic aspects of communication have intrigued game theorists, and there is a considerable body of literature on this topic by now. Most game theoretic studies of communication are not concerned with the specific properties of natural languages though. On the other hand, linguists have taken little notice of this line of research (with some notable exceptions like Arthur Merin and Prashant Parikh; cf. Merin 1999 or Parikh 2001) until the turn of the century, despite its obvious relevance. Within the last few years, this situation has changed somewhat. Various linguists and philosophers of language interested in pragmatics or language evolution started to study and employ game theoretic techniques. Research projects to this effect are under way in Amsterdam, Berlin, Bielefeld, at Northwestern University, the University of Pennsylvania, and perhaps at other places as well. A number of workshops about game theory and language/linguistics have taken place in recent years, organized both by linguists (like the special session on game theory at the Mathematics of Language meeting at Bloomington in 2003, the bi-annual conference ‘Games and Decisions in Pragmatics’, which takes place since 2003 in Berlin, or the colloquium ‘New Perspectives on Games

and Interaction' at the Royal Netherlands Academy of Arts and Science in 2007) and by economists (like a recent workshop on Communication, Game Theory, and Language at the Kellogg School of Business of Northwestern University). Last but not least, several biologists use the evolutionary interpretation of game theory to study the evolution of communication in biological systems, including natural language (see, for instance, the chapter on language evolution in Nowak 2006 and the references cited therein). In sum, a lively interdisciplinary community has emerged in recent years, which uses game theoretic techniques to study genuinely linguistic problems. The collection 'Game Theory and Pragmatics' (Benz et al. 2005), which contains contributions from linguists, economists, and philosophers of language, provides a representative sample of papers from this novel interdisciplinary field.

The topic of the present article is twofold. First, I will give a brief survey of main results in game theory about the nature of communication in general. Based on this, I will review some recent research on applications of the general framework to specifically linguistic problems.

There are two branches of game theory that are especially important for linguistics: 'standard' rationalistic game theory and evolutionary game theory. Because these two frameworks are conceptually quite different (even though they employ similar mathematical techniques), I will discuss them separately.

## 2 *What Is Game Theory?*

As mentioned in the beginning, game theory is a mathematical framework to study situations of strategic interaction. Such a situation always comprises at least two individuals. They are usually called the 'players'. Each player has choices between various ways of behaving – his strategies. Also, each player has 'preferences' over possible outcomes of the interaction. Preferences are represented by real numbers that are attached to each outcome of the situation (separately for each player), the 'utility'. Each player prefers higher over lower utilities. The outcome depends on the choices of each player. So in the general case, a given player cannot control that the situation develops in the way he prefers just by his own action – he has to take the actions of the other players into account. A situation with these properties is called a 'game'.

Decision-making can be non-deterministic. Mathematically speaking, the decision of a player does not have to be a function from situations to actions, but to probability distributions over actions. If such a distribution is non-trivial (assigns positive probability to several strategies), we speak of 'mixed strategies'. This can be interpreted epistemically (the player is not fully resolved what strategy to play) or frequentistic (the same player acts differently at different occasions even though the setting is identical, or identical players act differently in comparable situations).

In the simplest case, each player has one decision to make, and all players make their decision simultaneously, without knowledge about the actions of the other players. A typical example would be the Rock, Paper, Scissors<sup>1</sup> game. More complex games involve a sequence of alternating moves, where each player has complete or partial knowledge about the history of the game. Chess would be a case in point – a boring one though from the game theoretic point of view, because it can be proved that there are optimal strategies for both players; we just happen not to know what they are, because the search space is too large. A more interesting case is poker, where the players have only partial knowledge about the state of the game in most stages.

The mentioned games are ‘zero-sum games’, because ‘preferences’ can be represented as ‘utilities’ (like 1 for winning, –1 for losing, and 0 for a draw) in such a way that the utilities of all players always sum up to 0. Game theory can also model situations where the interests of the players partially or completely coincide (cf. Schelling 1960). A standard example is the following: suppose you want to dine out with a friend. Neither of you really cares about the restaurant; you simply prefer dining together over dining alone. Then the strategies are the restaurants under consideration. If both of you go to the same restaurant, you both obtain a utility of 1, otherwise you both get 0. Here the challenge is not to outsmart your co-players but to coordinate. Since here the interests of the players completely coincide, this is called a ‘partnership game’.

Game theory aims to predict how agents behave in such situations. Under the standard interpretation, these are normative predictions. Game theory can be used to develop recipes how a perfectly rational player ought to behave in a game (provided it is common knowledge that all players are perfectly rational). A perfectly rational player is an agent who strives to maximize his expected utility (given a certain, perhaps partial, knowledge about the situation and the other players) and who is logically omniscient. While this notion of rationality is an idealization, it is a useful approximation in many cases.

The most basic solution concept of standard game theory is ‘rationalizability’. A strategy is rationalizable if there is a consistent belief state of the player in question that makes this strategy rational. A more restrictive concept is the notion of a Nash equilibrium. This is a strategy profile (i.e. an assignment of a strategy to each player) where each player plays a best reply to the strategies of the other players. Another way of putting it is to say that in a Nash equilibrium, each player believes that the other players know his (possibly mixed) strategy in advance. In Rock, Paper, Scissors, there is just one Nash equilibrium – each player plays each strategy with equal probability. In the ‘partnership game’ mentioned above, there are several Nash equilibria: both of you going to restaurant A, both going to restaurant B, etc. Next to these desirable equilibria, there are less enjoyable ones as well though: if you play a mixed strategy that assigns the same

probability to each restaurant, then your partner will have no reason to prefer one restaurant over another, and she might play the same mixed strategy (which would leave you without any incentive to prefer a particular restaurant, so this is in fact a Nash equilibrium).

In 1973, the biologists John Maynard Smith and George Price published an article where game theory is used to study Darwinian natural selection. This article sparked the development of an entire branch of biomathematics, evolutionary game theory. The mathematical language is very similar to the one of standard game theory, but the interpretation is quite different. Instead of single agents, we are now dealing with populations of individuals. These individuals interact (both with members of the same population and with members of different populations), and the outcomes of these interactions affect reproductive success. The utility that each participant of such an interaction obtains is the effect of the interaction on his fitness, that is, the expected number of offspring. Also, the players do not 'decide' (in any useful sense of the term) which strategy to play. Rather, strategies are interpreted as genetically determined dispositions how to behave. If your genetic disposition ensures you a high utility in many interaction situations, your fitness will be high and you will have many offspring that inherit this disposition. So in general, strategies that lead to high utility will spread in the population (and vice versa – strategies that do poorly will die out).

In the population dynamic (or evolutionary) interpretation, a mixed strategy is a state of a population where different strategies are present. So each individual is disposed to act in a deterministic way, but the population as a whole acts non-deterministically.

Up to a point, there is a neat correspondence between rationalistic and evolutionary solution concepts. It is easy to show that all strategies that are not rationalizable under the standard interpretation will die out in the long run under the evolutionary interpretation. Also, if a population (or a configuration of populations in a multi-population setting) is in a Nash equilibrium state, it will stay there forever unless external forces act on the system.

Maynard Smith and Price proposed a solution concept that is somewhat stronger than the notion of a Nash equilibrium though. A state of the population is evolutionarily stable if it is a Nash equilibrium, and the state is furthermore protected against the occurrence of a small amount of mutation, that is, unfaithful reproduction.

To illustrate the notion of evolutionary stability, consider again the coordination problem where you want to meet somebody at a restaurant. The evolutionary interpretation of game theory can straightforwardly be applied to cultural phenomena as well (like the choice of a restaurant). Instead of biological reproduction, the replication of strategies proceeds via imitation (including self-imitation) in the cultural sphere. The utility of a strategy can be identified with its likelihood to be imitated.

Imagine you always have lunch in one of two restaurants A and B. Occasionally, you see another person at lunch who apparently also likes these two places. If you like this person, this may be a (perhaps small) incentive for you to go to the place where you met that person before. Suppose that person also enjoys seeing you at lunch. Then there are three Nash equilibria: both of you always go to A, you both always go to B, or you both always go to either place with equal probability. The first two states are evolutionarily stable. If you always meet each other at A, you will continue to go there even if your friend fails to show up once, and the same holds if you always go to B. Now suppose though that you both go to either place with equal probability (and in stochastically independent way). Then, due to sampling effects, you will encounter each other more often in A than in B (or vice versa). This provides an incentive to go even more often to A (B). This bias is self-reinforcing, and after some time you will end up in one of the two evolutionarily stable states.

### 3 Rationalizable Communication

Communication can be considered a game, but it can also be considered as an extension of an existing game. Reconsider the partnership game described in the last section where two people want to coordinate on a certain restaurant. The simplest way the players have to reach an optimal outcome is to talk to each other in advance and to agree on a place to meet. Communication will not alter the way zero-sum games are played though. In Rock, Paper, Scissors, it is advisable not to talk in advance, or at least not to reveal one's own plans.

A fair amount of the game theoretic literature on communication deals with issues like 'When is it rationalizable to communicate at all?' and 'How can credibility of messages be established?' One way to establish credibility is 'costly signalling'. Suppose you want to signal to the people around you that you are wealthy. Simply saying 'I am wealthy' will not be credible because poor people also prefer to be perceived as rich and may claim to be so. So cheap talk, that is, verbal boasting, will not help. A more convincing way to get the message across is to display your wealth, for instance, by driving a Jaguar. A poor man cannot use this signal simply because it is too expensive. So conspicuous consumption, or costs of signaling in general, may help to establish credibility.

However, if the interests of sender and receiver are sufficiently aligned, even cheap talk (meaning: communication where the signaling as such does not influence the utility of the players) can be credible. A nice example to illustrate this point comes from Rabin (1990). Suppose there are two players, the sender S and the receiver R. The sender belongs to one of the three types,  $t_1$ ,  $t_2$  or  $t_3$ . S knows her type, but R does not, and R considers all three types to be equally likely. R has a choice between three actions,  $a_1$ ,  $a_2$  and  $a_3$ . The utility of S and R both depend on S's

**Table 1. Utility matrix.**

	$a_1$	$a_2$	$a_3$
$t_1$	10, 10	0, 0	0, 0
$t_2$	0, 0	10, 10	5, 7
$t_3$	0, 0	10, 0	5, 7

type and on R's action. (You can imagine S being a job applicant and R her employer, and the three types are levels of skill and education. R's action might be different jobs that he can assign to S.) These utilities are represented by Table 1. Rows represent S's types and columns R's actions. The first number in each cell gives S's utility, and the second one R's utility.

Suppose S can send a message to R before R takes his action. S and R speak the same language, which is rich enough to express every proposition of the form 'I am T', where T is any non-empty set of types. If both players are rational, what kind of communication will ensue?

If S is  $t_1$ , she wants R to take action  $a_1$ , because this will maximize her payoff. If R believes that S has type  $t_1$ , he will in fact take action  $a_1$ , because this will maximize his own payoff. So in  $t_1$ , S has good reason to reveal her type by saying 'I am  $t_1$ .' If S is  $t_2$  or  $t_3$ , S would want R to take the action  $a_2$ . R would not do this if he believed S to be  $t_1$ . So if S is not  $t_1$ , she has no reason to pretend to be  $t_1$ . Hence, the message 'I am  $t_1$ ' is entirely credible – S wants R to believe that she is  $t_1$  if and only if she is  $t_1$ .

Things are different if S is  $t_2$ . In this case, she wants R to take action  $a_2$ , which R would only do if he is convinced that S is in fact  $t_2$ . So prima facie the message 'I am  $t_2$ ' is also credible. However, if S were  $t_3$ , she would also want R to take action  $a_2$ , and the only way to manipulate R in doing so would be to pretend that she is  $t_2$ . So the message 'I am  $t_2$ ' is not credible after all, because there are situations where S wants it to be believed even though it is not true.

The message 'I am  $t_3$ ' is certainly not credible, because it is never in the interest of S to make R believe it. If R believed S to be  $t_3$ , he would take action  $a_3$ , which is sub-optimal for S.

S can also send weaker messages. If she said: 'I am either  $t_2$  or  $t_3$ ', this would be credible again. If R believes it, he will take action  $a_3$  (which gives him an expected payoff of 7, while  $a_2$  would, on average, only give him 5). S would certainly prefer R to take action  $a_2$  in these cases, but the argumentation above showed that there is no way for her to manipulate R into taking this action.  $a_3$  is still better for S than  $a_1$  if the message is true, so S does have an interest in conveying that message if it is true, and she has no such interest if it is not true.

There is no credible message that is consistent with  $t_1$  except the most explicit 'I am  $t_1$ .' So the only two rational messages in this game are 'I am  $t_1$ ' and 'I am  $t_2$  or  $t_3$ '.

For game theorists, this reasoning still hides a puzzle. If S says 'I am  $t_1$ ' if she is  $t_1$ , and 'I am  $t_2$  or  $t_3$ ' if she is  $t_2$  or  $t_3$ , and R always believes her, this would constitute a Nash equilibrium. However, if S sends 'I am  $t_2$  or  $t_3$ ' in  $t_1$  and 'I am  $t_1$ ' if she is not  $t_1$ , and R always believes the negation of the literal meaning, this would also be a Nash equilibrium. In linguistic terms, using sentences in their literal meaning is exactly as rational as using them ironically, provided the conveyed messages are credible and R understands them correctly. Rabin proposes that the former equilibrium is the one that people will prefer, because it is focal. Briefly put, Rabin assumes that people will tell the truth unless they have a good reason not to. His formal notion 'credible message rationalizable strategies' captures the intuition that rational players will say/believe the truth as long as it is in their rational interest. (Similar ideas have been developed by other authors as well; see, for instance, Farrell 1993.)

Rabin's approach is reminiscent of Gricean reasoning.<sup>2</sup> Grice's overarching 'Cooperativity principle' can be translated into the game theoretic assumption that S and R always play partnership games where their interests always coincide. The preference, everything else being equal, for saying the truth is captured in the maxim of quality. Unlike Rabin, Grice considers situations, though, where talk is not necessarily cheap. The maxim of manner essentially says that S and R prefer short and concise over long and convoluted messages. This can be translated into the game theoretic language by assigning costs to signals.

If we assume cooperativity, every message is credible. Nevertheless, Gricean pragmatics teaches us that it is not always rational to use natural language expressions in their literal interpretation. Let us consider a very simple example to illustrate how the assumptions of rationality of interlocutors, plus a *ceteris paribus* preference for honesty, goes a long way to reconstruct the Gricean mechanism of implicature computation. Suppose R is planning to host a party, and he is asking S how many of the girls are going to show up. For simplicity, we only consider three options ('types' in game theoretic parlance, or 'possible worlds', as model-theoretic semanticists would say):

- $t_1$ : no girl comes to the party;
- $t_2$ : some but not all girls come to the party; and
- $t_3$ : all girls come to the party (and there are girls).

As far as R knows, all three types are equally likely. S knows the full truth. Both S and R have an interest that R gets as much information as possible (which corresponds to the first part of the maxim of quantity). This can be captured by the idea that the posterior probability that R assigns to the true type (after receiving S's message) is added to the utility

of both players. If R learns the true type, this will be 1. If he gets wrong information, this is 0. If he gets partial information, this value is  $1/3$  (if all three types remain possible) or  $1/2$  (if one wrong type can be ruled out). S can send either of four messages:

- $f_1$ : some girls will come to the party;
- $f_2$ : all girls will come to the party;
- $f_3$ : no girl will come to the party; and
- $f_4$ : some but not all girls will come to the party.

$f_1$ ,  $f_2$ , and  $f_3$  are approximately equally complex, while  $f_4$  is more complex. So according to the maxim of manner, S and R prefer, everything else being equal, not to use  $f_4$ . Let us say that using  $f_4$  reduces both player's utility by some small amount.

In  $t_1$ , S can only send  $f_3$  truthfully, and R has no reason to doubt its truth. In  $t_3$ , S could send  $f_1$  or  $f_2$ . If she sends  $f_2$  and R believes her, both obtain the maximal payoff. Sending  $f_1$  might lead S to assign some probability mass to  $t_2$ , so the payoff would not be maximal. Hence, the only rational action of S in  $t_3$  (provided she is not 100% sure which strategy R will play) is to send  $f_2$ . In  $t_2$ , S might send either  $f_1$  or  $f_4$ . R can infer from the previous considerations that  $f_1$  will never be sent in  $t_3$ . So in either case, R will assume correctly that  $t_2$  is the case. Since  $f_1$  is less costly than  $f_4$ , it is rational to send  $f_2$  in  $t_3$ . So if both players are rational and assume the maxim of quality, the interpretation of 'some' will be strengthened to 'some but not all', because if all girls came to the party, S had used the more specific expression  $f_2$ .<sup>3</sup>

The discussion of this example did not deal with the second part of the maxim of quantity (Do not provide more information than necessary!) and the maxim of relevance (Be relevant!). How relevance can be incorporated into the game theoretic set-up is, for instance, discussed in Benz and van Rooij 2007 and van Rooij 2003.

#### 4 Language Evolution

As mentioned in the beginning, evolutionary game theory is a reinterpretation of game theory that can be used to model evolutionary processes. The objects of study are not rational agents but evolving populations of entities, different strategies correspond to different traits of members of the population, and utility is to be interpreted as expected replicative success. The framework thus models Darwinian natural selection. More specifically, it models 'frequency dependent selection'. This means that the frequency distribution of strategies/traits within the population can affect the fitness of each individual. Evolutionary game theory is not so much concerned with the actual evolutionary dynamics of such a system. Rather, it studies the conditions for 'evolutionary stability'. Briefly put, the stable states are those states that are attainable and – up to a point – resistant against change.

Evolution and selection will always take place if a population of entities has some key properties:

- the elements of the population replicate;
- certain features of individuals are passed on faithfully from ‘parents’ to ‘offspring’;
- there is variation, that is, individuals within the population differ with respect to these heritable traits; and
- the expected number of offspring of an individual is correlated with its heritable traits.

These features are not confined to populations of living organisms. In particular, the elements of such a population can be behavioural traits of humans, or actual behavioural acts. Both types of entities get replicated via imitation. Therefore, evolution via selection takes place in the cultural sphere as well. This has been noticed among others by economists, and evolutionary game theory (which was developed by biologists that got some inspiration from economics) inspired economic research since the 1990s.<sup>4</sup>

The literature on evolutionary game theory contains some general results about stability of communication games that I will review now.

The simplest kind of communication game is a cheap talk signalling game. This means that there are (again) two players, sender *S* and receiver *R*. The sender has private information about some type or event *t*, which belongs to some finite set of events *T*. *R* does not know about this event, and it is in the interest of both players that *S* shares her knowledge with *R*. *S* can transmit one out of a finite number of signals *F*. So a strategy for *S* is a mapping from *T* to *F*. After observing the signal, *R* makes a guess about the identity of the event. Hence, a strategy for *R* is a mapping from *F* to *T*. If *R* guesses the event correctly, both players obtain a positive utility (which may differ between events); otherwise their utility is 0. Wärneryd (1993) and, in a more general setting, Trapa and Nowak (2000) show that such a game has an evolutionarily stable state if and only if the number of events equals the number of signals. In this case, the evolutionarily stable states (ESS) are exactly those strategy combinations where *S*’s strategy is a one–one map from events to signals, and *R*’s strategy is the inverse of this map.

This result indicates that natural selection alone is sufficient to establish a reliable communication system. It only applies though if the number of signals and events coincide. It is quite straightforward, however, to generalize it. Suppose there are more signals than events. Then the system will necessarily evolve towards a state where communication is perfect. In such a state, there may be synonymous signals (i.e. *S* might use a mixed strategy where the same event is mapped to several signals with a certain positive probability, and where *R* interprets each signal correctly). Alternatively, some signals may remain unused in such a state.

Such a state is, technically speaking, not an ESS, because it is not protected against drift. Suppose there are two events,  $t_1$  and  $t_2$ , and three forms,  $f_1$ ,  $f_2$ , and  $f_3$ . Let S use the strategy where  $t_1$  is always expressed as  $f_1$ , and  $t_2$  is expressed as  $f_2$  or  $f_3$  with a probability of 50% each. R interprets  $f_1$  as  $t_1$ ,  $f_2$  as  $t_2$ , and  $f_3$  also as  $t_2$ . So each event is reliably communicated. Now suppose due to some random effect, S changes her strategy to 50.1% probability to map  $t_2$  to  $f_2$ , and only 49.9% to map  $t_2$  to  $f_3$ . Since this also ensures maximal utility for both players (provided R does not change his strategy), natural selection will not push the system back to the original state. This, however, would be necessary for the original state to be an ESS. The set of states where communication is optimal forms an evolutionarily stable set (Thomas 1985) though. This is a set of strategy combinations with the property that the system will not leave that set once it is attained (provided external forces are sufficiently small).<sup>5</sup> If there are fewer signals than events, evolution will lead to a state where R's strategy is a one-one map (i.e. there are no synonymous signals) and S's strategy is consistent with the inverse of R's strategy. Furthermore, if the number of signals is  $n$ , exactly the  $n$  most important events will be expressible in the evolutionarily stable state. Importance of an event  $t$  is identified with the utility that the players obtain if  $t$  is correctly communicated, multiplied with the probability of  $t$ .<sup>6</sup>

Nowak et al. (1999) study a refined model, which has perhaps more significance for linguistics. Here, signals are not just atomic entities, but are located in some bounded metric space. You can think of signals as phonetic events with some articulatory or acoustic parameters, like voice onset time of a plosive or the formant frequencies of a vowel. Signal transmission is subject to some noise, so R might perceive a different signal than the one that S intended to transmit. The probability of such an error is the higher the closer the two signals are within the metric space. The authors show that under these conditions, there is always an upper limit for the number of different events that can be communicated. If the number of signals exceeds a certain threshold, the expressiveness that is gained by introducing a new signal is balanced by the loss in accuracy due to transmission errors. This is suggestive for the phonetics–phonology interface because in all languages, continuous phonetic variables are mapped to discrete phonological contrasts. For instance, voice onset time is continuous, but phonology only at most employs three categories, voiced, voiceless, and aspirated.

Nowak et al. also point out that this bottleneck can be overcome by assigning meanings to strings of signals, rather than to atomic signals. In this way, fitness can be increased practically unboundedly. Linguistically speaking, this suggests that double articulation (the fact that the smallest units of linguistic form, phonemes, are smaller than the smallest meaningful units, morphemes) is the result of evolution.

In a pilot study, I used computer simulations to figure out what the evolutionarily stable states look like if the signal space has the space of the

two-dimensional vowel space.<sup>7</sup> The two dimensions are the first two formants. This two-dimensional space has roughly the shape of a triangle; the precise numerical model is taken from Liljencrants and Lindblom (1972). In each simulation, the number of vowel categories ('phonemes' if you like) remained fixed, and I ran simulations with this number ranging between 3 and 9. A state of the system consists of a memory consisting of a set of 1000 production events (pairs consisting of a vowel category and a point in the vowel space) and a set of 1000 perception events (pairs consisting of a phonetic point and a phonological category). In the beginning of a simulation, these pairings are randomized. A single round of the game proceeds as follows:

- a vowel category  $\nu$  is picked out at random;
- among all production events in memory that express this category  $\nu$ , one is picked out at random and the corresponding signal  $f_1$  is produced;
- a normally distributed random variable is added to this signal, resulting in signal  $f_2$ ;
- among all perception events  $\langle f', \nu' \rangle$ , the one is picked out which minimizes the distance between  $f_2$  and  $f'$ ; and
- if  $\nu = \nu'$ , the pair  $\langle \nu, f_1 \rangle$  is added to the memory of production events (replacing an older item), and the pair  $\langle f_2, \nu \rangle$  is added to the memory of perception events. Otherwise, the memory remains unchanged.

In this set-up, the evolving populations consist of phonetic events, not of language users. You can interpret this as an implementation of 'exemplar dynamics' in the sense of Pierrehumbert (2001).

Each simulation consists of 300,000 iterations. The results are depicted in Figure 1. Each colour represents a vowel category. The coloured dots

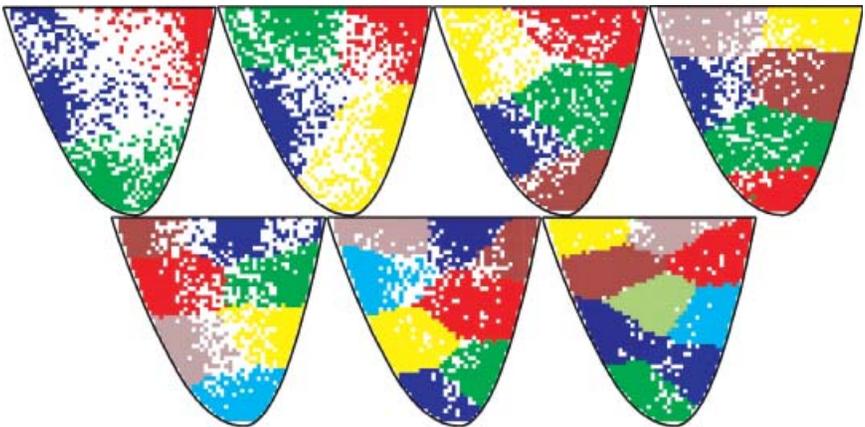


Fig. 1. Simulation results.

Number of vowels	Vowel systems and their frequency of occurrence				
3	 14				
4	 14	 5	 4	 2	
5	 97	 3			
6	 26	 12	 12	 4	
7	 23	 6	 5	 4	 3
8	 6	 3	 3	 2	
9	 7	 7	 3		

Fig. 2. Survey of vowel systems from Schwartz et al. (1997), published in *Journal of Phonetics* 25, The dispersion-focalization theory of vowel systems, 255–86, © Elsevier.

represent the receiver's strategy, that is, the category that the corresponding signal is most often mapped to. It is instructive to compare this with survey of vowel systems across the languages of the world from Schwartz et al. (which is based on the UCLA Phonetic Segment Inventory Database), given in Figure 2.

Of the 264 vowel systems covered by the survey, 181 more or less correspond to one of the configurations that came out as evolutionarily stable in the simulation. Also, six out of the seven outcomes of the simulations correspond to typologically common systems.

I hasten to add that these results are quite preliminary. The outcomes of simulations depend on a variety of parameters like the relative weighting of the two formants, the standard deviation of the noise variable, the

memory size, etc. In future research, it is planned to determine the stable states via analytic and numerical techniques, and to choose parameter values that are informed by phonetic research. Still, the fit between the initial simulation results and the typological findings is encouraging.

I would like to mention three further studies that use evolutionary game theory to explain linguistic universals and typological tendencies.

Van Rooij (2004) deals with the so-called 'Horn strategies'. By this he means the tendency of natural languages to pair simple forms with stereo-typical meanings, and complex forms with non-stereotypical meanings. While both a Horn strategy and its sub-optimal counterpart (pairing simple forms with complex meanings and vice versa) turn out to be evolutionarily stable, only Horn strategies are also stochastically stable. Stochastic stability is a refinement of evolutionary stability that takes sampling effects in finite populations into account (cf. Kandori et al. 1993 and Young 1993).

In Jäger (2007), the stability conditions for basic case-marking systems are investigated. The strategy space for the speaker consists of possible mappings from core syntactic roles (like agent or object) to case markings (nominative, accusative, or ergative), possibly conditioned by semantic prominence of the NP in question. Background assumptions are that nominative, being the unmarked case, is less costly than accusative or ergative, and that nominative is underspecified with respect to syntactic role, while accusative can only mark objects, and ergative agents. The utility function captures the idea that unambiguous encoding increases utility, while morphological complexity reduces it. The correlation between prominence of NPs and their syntactic roles that was determined via corpus studies was used as an invariant side condition. The evolutionarily stable states in this game comprise most of the typologically common case-marking systems, but also include some rare or unattested systems. Only the common systems turn out to be stochastically stable.

The basic set-up from Gärdenfors (2000) is taken up in Jäger and van Rooij 2007 (see also Jäger 2008). Gärdenfors assumes that meanings are arranged in conceptual spaces that have a geometrical structure. He argues at length that natural categories are convex regions of such a space. In the game theoretic reconstruction, the consequences are explored if there are many more meanings than words to refer to these meanings. In this scenario, perfect communication of all possible meanings is impossible. If meanings are arranged in a metrical space though, there are different degrees of miscommunication. The utility is the higher the smaller the distance is between what the sender wants to express and what the receiver actually perceives. With these assumptions, it turns out that evolution necessarily leads to a state where the sender strategy induces a Voronoi tessellation of the meaning space, which in turn entails the convexity of categories. A certain feature of categorization that Gärdenfors claimed to be rooted in cognition may thus turn out to

be social rather than cognitive in nature, because it emerges from the requirements of communication.

### 5 Conclusion

In this article, I gave a short survey of the state of the research in game theoretic linguistics. Both the rationalistic and the evolutionary branch of this paradigm are still in its infancy, and a lot of the current discussion revolves about rather basic issues how linguistic concepts are to be mapped to game theoretic ones. The most urgent issue is perhaps the question how utility functions are to be motivated. The applications that I reviewed above more or less borrow their preference orderings from older linguistic traditions like Gricean pragmatics or markedness theory. Strictly speaking, assumptions about a utility function are assumptions about decision-making of language users (in the rationalistic setting) or about the imitation likelihood of linguistic variables (in the evolutionary interpretation). Such assumptions can be tested psycholinguistically, and the combination of game theoretic and experimental methods is a very promising route for future research. Another open question is which solution concept is actually appropriate. Above I argued for iterated elimination of weakly dominated strategies (paired with a default preference for honesty) in the rationalistic setting. Parikh (2001), for instance, proposes Pareto-dominant Nash equilibria. Other proposals can be found in the literature as well. In the evolutionary setting, researchers have worked with evolutionarily stable states, evolutionarily stable sets, and stochastically stable states. Most likely, none of these solution concepts is appropriate for all applications. Rather, motivating a solution concept has to be part of an application of game theory, just like motivating a utility function. This will likely lead to an improved understanding of tacit modeling, and thus of the domain of application.

### Short Biography

Gerhard Jäger obtained his Ph.D. from Humboldt University at Berlin in 1996. After that he worked as a researcher at various universities in Europe and the USA, including Stanford, the University of Pennsylvania, and Utrecht University. Since 2004, he is professor of linguistics at Bielefeld University in Germany.

Jäger has worked on a variety of topics in theoretical and computational linguistics, ranging from dynamic semantics and categorial grammar to Optimality Theory. His current research interest is the (cultural) evolution of natural languages. In this context he works with mathematical techniques from theoretical biology and economics – especially evolutionary game theory, but also computer simulations. Together with Anton Benz and Robert van Rooij, he recently edited the volume *Game Theory and Pragmatics* (Palgrave Macmillan 2004).

## Notes

\* Correspondence address: Gerhard Jaeger, Department of Linguistics and Literature, University of Bielefeld, PF 10 01 31, Bielefeld, Germany. E-mail: gerhard.jaeger@uni-bielefeld.de.

<sup>1</sup> Rock, Paper, Scissors is a popular two-person game. In its best-known variant, each player has to choose between three options: rock, paper, or scissors, which are indicated by iconic gestures (a fist for rock, a flat hand for paper, and a gesture of the index finger and the middle finger similar to Churchill's victory sign for scissors). Both players make their choice simultaneously. Rock wins over scissors (because a rock blunts scissors), scissors wins over paper (because scissors cut paper), and paper wins over rock (because paper wraps the rock). If both players make the same choice, the outcome is a draw.

<sup>2</sup> See also Stalnaker (2005) on the relation between the game theoretic notion of credibility and Gricean pragmatics.

<sup>3</sup> The solution concept that has been used informally here is called 'iterated elimination of weakly dominated actions' in the literature. There is some debate about its general applicability, even though it does capture a plausible notion of rationality; see Chapter 6 of Osborne and Rubinstein (1994) for discussion.

<sup>4</sup> Language (both in the sense of *langue* and of *parole*) can be analysed as evolving entity. Of course, language, the human language faculty, is affected by biological evolution. So evolutionary theory can inform linguistics at all levels of language description.

<sup>5</sup> The fact that the optimal signalling systems form an evolutionarily stable set can be shown as follows: in partnership games, the average utility is a strict Lyapunov function (Hofbauer and Sigmund 1998). Therefore, the set of states where average utility attains its global maximum forms an asymptotically stable set of rest points. Cressman (2003) shows that in symmetrized asymmetric games, the ESSets are exactly the asymptotically stable sets of rest points. The maximal fitness is attained in the game in question if and only if each event is reliably communicated.

<sup>6</sup> This can be shown as follows: suppose otherwise, that is, (S,R) is a Nash equilibrium. Let there be two events,  $t_1$  and  $t_2$ , such that  $t_1$  is more important than  $t_2$ . Also, suppose that  $t_2$  is reliably communicated (every signal that  $t_2$  is mapped to with a positive probability under S is mapped to  $t_2$  with probability 1 under R), but  $t_1$  is not. This means there is some signal  $f$  such that S maps  $t_1$  to  $f$  with positive probability, and R maps  $f$  to  $t_1$  with a probability  $< 1$ . So R maps  $f$  to some event  $t^*$  with some positive probability. Hence, there is a best response  $R'$  to S that is like R except that it always maps  $f$  to  $t^*$ . So under (S,R'),  $t_1$  is never correctly communicated. Hence, there is a best reply  $S'$  to  $R'$  that is like S except that it maps  $t_1$  to some signal  $f^*$  that has the property that  $t_2$  is mapped to  $f^*$  with positive probability under S. Since  $t_1$  is more important than  $t_2$ , there is a best reply  $R''$  to  $R'$  that maps  $f^*$  to  $t_1$  with probability 1. The expected fitness of (S',R'') is higher than the fitness of (S',R'), and hence also as the fitness of (S,R). So (S,R) cannot be part of any set of asymmetric Nash equilibria that is closed under best response. According to Cressman (2003), the ESSets of some symmetrized asymmetric game are exactly the symmetrizations of strict equilibrium sets (SESet), and each SESet is a set of Nash equilibria that is closed under best response. Hence (S,R) cannot be part of any ESSet.

<sup>7</sup> The set-up of the simulation to be presented here is quite simplistic, compared with the sophisticated study of the evolution of vowel systems by de Boer (2001). The advantage of the simple model is that it lends itself to an explicit game theoretic analysis.

## Works Cited

- Benz, Anton, Gerhard Jäger, and Robert van Rooij (eds.). 2005. *Game theory and pragmatics*. Basingstoke, UK: Palgrave MacMillan.
- Benz, Anton, and Robert van Rooij. 2007. Optimal assertions and what they implicate. *Topoi* – An International Review of Philosophy 27.63–78.
- de Boer, Bart. 2001. *The origin of vowel systems*. Oxford, UK: Oxford University Press.
- Farrell, Joseph. 1993. Meaning and credibility in cheap-talk games. *Games and Economic Behavior* 5.514–31.

- Cressman, Ross. 2003. *Evolutionary dynamics and extensive form games*. Cambridge, MA: MIT Press.
- Gärdenfors, Peter. 2000. *Conceptual spaces*. Cambridge, MA: MIT Press.
- Hofbauer, Josef, and Karl Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge, UK: Cambridge University Press.
- Jäger, Gerhard. 2007. Evolutionary Game Theory and typology: a case study. *Language* 83.74–109.
- . 2008. The evolution of convex categories. *Linguistics and Philosophy*, forthcoming.
- Jäger, Gerhard, and Robert van Rooij. 2007. Language structure: psychological and social constraints. *Synthese* 159.99–130.
- Kandori, Michihiro, George Mailath, and Rafael Rob. 1993. Learning, mutation, and long-run equilibria in games. *Econometrica* 61.29–56.
- Lewis, David. 1969. *Convention*. Cambridge, MA: Harvard University Press.
- Liljencrants, Johan, and Björn Lindblom. 1972. Numerical simulations of vowel quality systems: the role of perceptual contrast. *Language* 48.839–62.
- Maynard Smith, John, and George R. Price. 1973. The logic of animal conflict. *Nature* 246.15–8.
- Merin, Arthur. 1999. Information, relevance, and social decisionmaking. *Logic, language, and computation*, vol. 2, ed. by Jonathan Ginzburg, Larry Moss and Maarten de Rijke. Stanford, CA: CSLI Publications.
- Nowak, Martin A. 2006. *Evolutionary dynamics. Exploring the equations of life*. Cambridge, MA: Harvard University Press.
- Nowak, Martin A., David C. Krakauer, and Andreas Dress. 1999. An error limit for the evolution of language. *Proceedings: Biological Sciences* 266.2131–6.
- Osborne, Martin J., and Ariel Rubinstein. 1994. *A course in game theory*. Cambridge, MA: MIT Press.
- Parikh, Prashant. 2001. *The use of language*. Stanford, CA: CSLI Publications.
- Pierrehumbert, Janet. 2001. Exemplar dynamics: word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul J. Hopper, 137–58. Philadelphia, PA: John Benjamins.
- Rabin, Matthew. 1990. Communication between rational agents. *Journal of Economic Theory* 51.144–70.
- van Rooij, Robert. 2003. Questioning to resolve decision problems. *Linguistics and Philosophy* 26.727–63.
- . 2004. Signaling games select Horn strategies. *Linguistics and Philosophy* 27.493–527.
- von Neumann, John, and Oskar Morgenstern. 1944. *Theory of games and economic behavior*. Princeton, NJ: John Wiley & Sons.
- Schelling, Thomas C. 1960. *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Schwartz, Jean-Luc, Louis-Jean Boe, Nathalie Vallé, and Christian Abry. 1997. The dispersion-focalization theory of vowel systems. *Journal of Phonetics* 25.255–86.
- Spence, Michael. 1973. Job market signaling. *The Quarterly Journal of Economics* 87.355–74.
- Stalnaker, Robert. 2005. Saying and meaning, cheap talk and credibility. *Game theory and pragmatics*, ed. by Anton Benz, Gerhard Jäger and Robert van Rooij, 83–100. Basingstoke, UK: Palgrave MacMillan.
- Thomas, Bernhard. 1985. On evolutionarily stable sets. *Journal of Mathematical Biology* 22.105–15.
- Trapa, Peter E., and Martin A. Nowak. 2000. Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology* 41.172–88.
- Wärneryd, Karl. 1993. Cheap talk, coordination and evolutionary stability. *Games and Economic Behavior* 5.532–46.
- Young, H. Peyton. 1993. The evolution of conventions. *Econometrica* 61.57–84.