Evolutionary Game Theory and Linguistic Typology: a Case Study

Gerhard Jäger*

Abstract
The paper deals with the typology of the case marking of semantic core roles. The competing economy considerations of hearer (disambiguation) and speaker (minimal effort) are formalized in terms of evolutionary game theory. It will be shown that the case marking patterns that are attested in the languages of the world are those that are evolutionary stable for different relative weightings of speaker economy and hearer economy, given the statistical patterns of language use that were extracted from corpora of naturally occurring conversations.

1 The frequencies of clause types

Consider all (logically) possible case marking types that only use case splits induced by the contrast between pronouns and full NPs. I will restrict attention to possible grammars where the morphological form of the intransitive subject (nominative/absolutive) is less complex than ergative and accusative (if present). Which language types are functional and which aren't? The main function of case marking is of course to disambiguate, i.e. to enable the hearer to identify the semantic role of the denotation of an NP. More particular, case should uniquely identify the argument roles "A" (agent, i.e. the transitive subject) and "O" (the direct object). We can assume without loss of generality that the hearer always interprets an ergative morpheme as A if there is one, and likewise an accusative morpheme as O, so ambiguity can safely be avoided if at least one NP per clause is case marked. For the sake of brevity, I will denote case marking patterns from now on as a quadruple of case forms, in the order: case of 1. pronominal agents, 2. non-pronominal agents, 3. pronominal objects, and 4. non-pronominal objects. Ergative marking is abbreviated as "*e*", accusative as "*a*", and zero marking (i.e. nominative/absolutive) as "*z*". For instance, a language like English where only pronominal objects are case marked would thus follow the pattern *zzaz*, while a language like Basque with obligatory ergative marking of all agents is *eezz*.

Ambiguity will only arise if a grammar admits clause types without any case marking. However, this need not lead to ambiguity if one of the two unmarked arguments is prominent and the other isn't. Then the hearer may em-

* Institute for Linguistics, University of Potsdam, e-mail: jaeger@ling.uni-potsdam.de

ploy a default rule to the effect that in such a case the more prominent NP is A (or vice versa). This taken into account, the speaker strategies *zeaz* and *ezza* also avoid ambiguity in the sense that there is a corresponding hearer strategy that always correctly identifies semantic roles. One might assume that word order is a good predictor of syntactic roles too, but even in languages with fixed word order there may occur elliptical expressions which are, without the aid of case morphology, ambiguous. Let us assume that disambiguation is the main priority of the speaker, but he has the secondary priority to use as few case morphemes as possible. It depends on the relative frequencies of clause types which patterns minimize the average number of case morphemes per clause. We only have to consider four clause types – both A and O may be p (pronominal) or n (non-pronominal). The percentages in figure 1 are extracted from Geoffrey Sampson's CHRISTINE corpus of spoken English, and I took pronouns to be prominent and full NPs to be non-prominent. The set of all clauses comprising a subject and a direct object amount to 100%.

	O/p	O/n
A/p	19.70%	71.24%
A/n	1.59%	7.46%

Figure 1: clause type frequencies

I will refer to the four cells of this table with pairs *pp*, *pn*, *np*, and *nn*, where the first element gives the specification of A and the second of O. The concrete figures of course depend on the corpus under investigation and the choice of the prominence split. However, for the results reported below, the only thing that matters is that $pn > np$, and this inequality robustly holds for all corpora (including both spoken and written corpora in English, German and Swedish) I investigated and for all split points along the definiteness hierarchy or the animacy (to the degree that the corpora investigated were annotated for animacy) hierarchy.

2 Game Theory

Game Theory is well-suited to make the possibly conflicting priorities of speakers and hearers more precise. Let us assume that a fixed set of meanings M and forms F is given. A speaker strategy is any function S from M to F, i.e. a production grammar. Likewise, a hearer strategy is a comprehension grammar, i.e. a function from F to M. In an utterance situation, the speaker has to decide what to say and how to say it. Only the latter decision is a matter of grammar; the decision about what meaning the speaker tries to communicate is related to other cognitive domains. Let us thus assume that in each game, nature presents the speaker with a meaning m, and the speaker only has to choose how to express m. Communication is successful if the hearer recovers the intended meaning from the observed form. It is measured by the δ-function:

$$\delta_m(S,H) = \begin{cases} 1 & \text{iff} \quad H(S(m)) = m \\ 0 & \text{else} \end{cases}$$

Forms differ with respect to their complexity. I take it that the complexity can be measured numerically, i.e. *cost* is a function from F to the non-negative real numbers. The speaker has two possibly conflicting interests: he wants to communicate the meaning as accurately as possible while simultaneously minimizing the complexity of the form used. This is captured by the following definition of speaker utility:

$$u_s(m, S, H) = \delta_m(S, H) - k \times cost(S(m))$$

Here k is some positive coefficient that formalizes the priorities of the speaker. A low value for k means that communicative success is more important than minimal effort and vice versa. The hearer tries to recover the intended meaning as accurately as possible. So the hearer utility can be identified with the δ-function:

$$u_h(m, S, H) = \delta_m(S, H)$$

Nature presents meanings to the speaker according to a certain probability distribution x. The average utilities of speaker and hearer in a game can thus be given as

$$\begin{aligned} u_s(S, H) &= \sum_m x_m \times (\delta_m(S, H) - k \times cost(S(m))) \\ u_h(S, H) &= \sum_m x_m \times \delta_m(S, H) \end{aligned}$$

We are only concerned with elementary transitive clauses. So we are dealing with two NPs. One is A and the other O, and both may be either p or n. I am not concerned with the effect of word order or head marking on argument linking in this paper. Therefore I take it that nature chooses the word orders $A - O$ and $O - A$ with a 50% probability each, and that this choice is stochastically independent from the specifications of the NPs as p or n. Furthermore nature specifies which of the two NPs is A and which is O, and whether they are n or p. This gives a total of eight meanings. x is a probability distribution over these eight meanings. It is plausible to assume that the prominence of an NP is always unambiguously encoded in its form. This leaves us with 36 possible forms — each of the two NPs may be p or n, and either one may be marked with ergative, accusative, or zero case. The cost function simply counts the number of case morphemes per clause.

I will restrict attention to just a small subset of simple strategies. First, word order effects are kept out of considerations. Furthermore, I take it that the case morphology of a given NP only depends on its own prominence value and syntactic function, not on the prominence value of the other NP. Among these strategies, I will restrict attention to those where the two marked forms are reserved for one syntactic role each while the unmarked form is in principle ambiguous between A and O. This leaves us, modulo renaming of e and a, with 16 case marking patterns, $eeee, eeaz, eeza, \cdots, zzza, zzzz$. Of these 16 strategies, 6 are strictly dominated (i.e. they are never optimal, no matter what the hearer does), namely those that sometimes use two case morphemes per clause, and the inverse split ergative pattern *ezza*.

A hearer strategy is a mapping from forms to meanings. If ergative is only used to mark A by the speaker and accusative only for O, it would obviously be unreasonable by the hearer to interpret the case morphemes otherwise. I will call the hearer strategies that interpret ergative as A and accusative as O "faithful." There are only 16 faithful strategies. Thus only the interpretation of clauses without case morphology is undetermined. There are four such clause types (depending on the prominence features of the two NPs), each of which may receive two possible interpretations. If both NPs in a form f have the same prominence value, both interpretation strategy classes have actually the same expected payoff because by assumption, the speaker strategies exclude correlations between word order and meaning, and the prominence values give no clue. So we may safely identify any pair of hearer strategies which only differ in their interpretation of $p/z - p/z$ or $n/z - n/z$. Now we are down to four hearer strategies — they differ with respect to the meaning they assign to $p/z - n/z$ and $n/z - p/z$. I will denote these strategies as AA, AO, OA and OO, where the first component is the interpretation of the first NP in $p/z - n/z$, and the second component the interpretation of the first component of $n/z - p/z$.

The configuration of Nash Equilibria (NEs henceforth) depends on the value of k. For small values of k, the split ergative pattern $zeaz/AO$ is a strict NE (i.e. each component strategy is the **unique** best response to the opponent's strategy). Besides, each combination of a pure ergative (*eezz*) or pure accusative (*zzaa*) speaker strategy with any hearer strategy $\neq AO$ is a non-strict NE. For larger values of k, two strict NEs coexist, either differential object marking ($zzaz/AO$) and inverse differential subject marking ($ezzz/OA$), or differential subject marking ($zezz/AO$) and inverse differential object marking ($zzza/OA$). Finally, for very large values of k, the system without case marking $zzzz/AO$ is the unique (and hence also strict) NE.

Let us take stock. Of the sixteen case marking strategies that we considered, only eight give rise to an NE in some configuration. The eight strategies that were excluded are in fact typologically unattested or at least very rare. There is apparently only one language with a full-blown tripartite system, i.e. with the strategy *eeaa*, namely the Australian language Wangkumara. Inverse split ergative systems — *ezza* in my notation — are also very rare. It is a bit tricky to decide whether languages of the type *zeaa* or the like exist. There are several split ergative languages where the split points for ergative and accusative differ, and where there is an overlap in the middle of the hierarchy with a tripartite paradigm. Since the system I use here implicitly assumes that the two split points always coincide, such languages cannot really be accommodated; they are a mixture of *eeaz*, *zeaa* and *zeaz*. To my knowledge, clearcut instances of *eeaz* or *zeaa* do not exist, and the combinations *ezaa* and *eeza* are unattested as well. There are no languages which would have a tripartite paradigm for all and only the prominent or all and only the non-prominent NPs.

Hence *zeza* and *ezaz* are correctly excluded. So the concept of a Nash Equilibrium proves fairly successful in identifying possible case marking systems. Conversely, we expect to find instances of languages with an NE pattern. This is certainly the case for *zzaz* (like English), *zezz* (for instance the Circassian languages Adyghe and Kabardian), *zeaz* (like Dyirbal), and *zzzz* (as in several Bantu languages). However, the concept is still too inclusive. I know of only one language of the types *zzza* and *ezzz* each, namely Nganasan (see [3], p. 90) as instance of the former and (according to [1]) Wakhi of the latter. The pure accusative systems — *eezz* — do exists (Hungarian is an example), but they are also very rare. Most accusative languages have DOM, and most ergative languages DSM. Besides, the rationalistic approach has the same conceptual problem as any functional explanation of grammatical patterns: natural languages are not consciously designed, and it is *a priori* not clear at all why we should expect to find functionally plausible patterns.

3 Evolutionary Game Theory

In Evolutionary Game Theory (EGT), we are dealing with populations of players that are programmed for a certain strategy. Players replicate and pass on their strategy to their offsprings. The number of offsprings is directly related to the average payoff of the parent strategy.

How can this model be applied to linguistics? If the strategies in the EGT sense are identified with grammars (as done in the previous section), games should be identified with utterances. However, grammars are not transmitted via genetic but via cultural inheritance. Therefore, **imitation dynamics** is more appropriate here than the replicator dynamics that is used in applications of EGT to theoretical biology. According to the imitation dynamics, players are not mortal and have no offsprings. However, every so often, a player is offered the opportunity to pick out some other player and to change his own strategy against the strategy of the other player. The probability that a certain strategy is adopted for imitation is positively correlated to the gain in average utility that is to be expected by this strategy change. So here as well as in the standard model, successful strategies will tend to spread while unsuccessful strategies die out. Moreover, exactly the same strategies are evolutionary stable under the replicator dynamics and under the imitation dynamics. Several sources of mutations are conceivable here, ranging from plain speech errors to socio-linguistic factors like language contact. We expect that most natural language grammars are evolutionary stable because unstable grammars do not persist. The Game of Case that was introduced in the last section is an **asymmetric game**. In a population dynamic setting, this means that we are dealing with two separate populations. So rather than with evolutionary stable strategies, we have to deal with evolutionary stable strategy pairs here. In multi-population dynamics, evolutionary stability can be characterized quite easily in rationalistic terms. Briefly put, a strategy pair is evolutionary stable iff it is a **Strict Nash Equilibrium** (SNE henceforth).

Let us apply the analytical tools of EGT to the different instantiations of the Game of Case. The NEs using a pure case marking strategy (*eezz* or *zzaa*) are never strict and thus not evolutionary stable. The remaining 6 NEs are strict though. Of these 6 strategy pairs, two are very rare among the languages of the world: *zzza*/*OA* and *ezzz*/*OA*. Put differently, it is important to note that these two "wrong" SNE each coexist with a well-attested SNE, namely inverse differential subject marking (*ezzz*/*OA*) with differential object marking (*zzaz*/*AO*), and inverse differential object marking *zzza*/*OA* with differential subject marking *zezz*/*AO*. In both scenarios, the typologically attested SNE is Pareto-optimal, i.e. it has a higher average utility than the competing SNE.

The standard approach to EGT assumes that populations are infinite. If we assume instead that the populations are finite but large, every invasion barrier is occasionally broken, no matter how low the mutation rate is. (With increasing population size, the likelihood of such an incident converges towards 0.) It can be shown that the Pareto-optimal SNE always has a higher invasion barrier than the other SNE. In a finite population, it is thus more probable to switch **to** than **from** the Pareto-optimal SNE. In a *population of finite populations*, the unique attractor state is the one where the majority of population is in the Pareto-optimal SNE, and as the size of the single finite populations increases, the probability of the non-Pareto-optimal SNE converges towards 0. (See [4] for a similar explanation of the asymmetry between multiple evolutionary stable strategies.)

To sum up, under the assumption of a population of finite but large populations of speakers/hearers, only four strategies are evolutionary stable: split ergativity, differential subject marking, differential object marking, and absence of case marking. This fits the typological findings rather well. While the majority of languages is in an evolutionary stable state, there are some exceptions. Evolutionary Game Theory predicts that such language types should be diachronically unstable. This is an empirically testable claim that should be tackled in future research.

References

[1] Elena Bashir. Beyond split-ergativity: Subject marking in Wakhi. In *Papers from the 22nd regional meeting of the Chicago Linguistic Society*, pages 14–35. CLS, Chicago, 1986.

[2] Barry J. Blake. *Case*. Cambridge University Press, 2001.

[3] R. M. W. Dixon. *Ergativity*. Cambridge University Press, Cambridge, 1994.

[4] Robert van Rooy. Signaling games select Horn strategies. manuscript, University of Amsterdam, 2002.