Evolutionary Game Theory and Typology: a Case Study

Gerhard Jäger

University of Bielefeld

Gerhard.Jaeger@uni-bielefeld.de

address: University of Bielefeld, Dept. of Linguistics and Literature, PF 10 01 31, D-33501 Bielefeld, Germany

Evolutionary Game Theory and Typology: a Case Study

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 is shown that the case marking patterns that are attested in the languages of the world are those that are evolutionarily 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.*

*The research that led to this paper has been sparked by many discussions with colleagues mainly from Amsterdam, whose thoughts went along similar lines as mine at that time. It is sometimes hard to determine in retrospect who had what idea first, but I will at least attempt to give the due credits. Henk Zeevat was the first who suggested to me that the typology of case-marking patterns might be related to utterance frequencies, and that this hypothesis should be validated with the help of corpora. This program has been spelled out in a different framework in Zeevat and Jäger 2002. Similar ideas had been pursued around the same time by Marie Nilsenova (see Nilsenova 2002). William Rose picked up the idea as well and also formalized it using Evolutionary Game Theory. The details of his formalization differ from mine, but there is a considerable overlap. Especially the insight that the split ergative system is, under certain side conditions, the only evolutionarily stable strategy has been reached by Rose (in still unpublished work) much earlier than by me. Judith Aissen and Joan Bresnan's course on Typology and Optimality Theory at the Düsseldorf Summer School 2002 provided a plethora of insights and inspirations. Robert van Rooij's pioneering work on the application of EGT to linguistics (van Rooij 2004), as well as ongoing discussions with him, were important for me to implement the EGT formalization.

Furthermore, Manfred Krifka, Anette Rosenbach, the editors of Language and three anonymous referees made a lot of insightful comments on a draft version of this paper. Last but not least I have to thank the audiences of related talks in Amsterdam, Berlin, Osnabrück and Potsdam for useful feedback, and many colleagues for discussions, among them Judith Aissen, David Beaver, Peter beim Graben, Reinhard Blutner, Paul Boersma, Joan Bresnan, Helen de Hoop, Paul Kiparsky, Lena Maslova, Åshild Næss, Barbara Stiebels, and Henk Zeevat.

1 Introduction The basic function of case-marking is arguably the identification of syntactic core roles. Among all logically conceivable case-marking systems, a surprisingly small number occurs very often among the languages of the world. These are accusative systems with differential object marking, ergative systems with differential subject marking, and combinations of these two paradigms. Other patterns are either very rare or completely unattested. A good case can be made that the systems that are frequent are exactly those that are well-adapted to their function in language use.[1] Drawing on an analogy from biology, this kind of functional adaptation might be the result of shaping by the forces of evolution.

Evolutionary approaches to linguistic problems have received growing attention from different quarters of the scientific community in past years. The issues of the biological origin of natural language and the human language faculty are obviously related to human evolution, but evolutionary thinking has been fruitfully applied to the development of natural languages on a historical time scale as well. Languages or grammars can be seen as self-replicating systems, and as soon as replication is subject to variation and selection, evolutionary processes will emerge. Evolutionary linguistics is still in a 'pre-paradigmatic stage' (Jelle Zuidema, p.c.) though, and there is no agreement so far about such fundamental questions as the units of replication, the source of variation and the criteria of selection.

In the study presented here, I use an evolutionary approach to explain the said patterns in the typology of case-marking across the languages of the world. Before starting a discussion of the empirical issues to be investigated, let me briefly sketch the foundational and methodological decisions that underlie this work.

It seems plausible to assume that functional, cognitive and social factors interact in linguistic selection. I focus here solely on functional factors—grammars that increase communicative success and minimize the speaker effort are considered 'fitter' than competing grammars that are less successful in this respect. The decision to concentrate on these aspects is partially a pragmatic one—the communicative function of a linguistic sign is much easier to model than the gain in social prestige that comes with its usage, say.

Evolutionary concepts have recently been discussed in historical linguistics (see for instance

Haspelmath 1999 or Croft 2000). My own focus is less on language change though. It is important, I would say, to ask why certain features of languages are typologically common and stable than to ask why and how other features change. This point can be illustrated with an example from biology. Fisher (1930) investigated the evolutionary dynamics of the different possible sex ratios in sexually reproducing species. He was able to show that a sex ratio of 1:1 is the only evolutionarily stable option. This is in fact a good approximation of what we find across the species of the animal kingdom, and for all we know, this feature hasn't changed for the past millions of years. In this case evolutionary thinking gives a profound explanation for an entirely static biological fact.

Finally, my analysis is couched in terms of Evolutionary Game Theory (EGT). This mathematical framework to model evolutionary processes is the result of joint efforts of theoretical biologists and economists. I discuss the details and its applicability to linguistic problems later on, but I mention a methodological aspect here. EGT is a sophisticated theoretical tool, and there are many 'off-the-shelf' methods and results from this area that can be applied right away if the empirical domain to be modeled is formulated in the right way. It strikes me as worthwhile to use these tools rather than starting from scratch.

2 Differential Case Marking Most languages of the world have mixed case-marking patterns (see for instance Bossong 1985; detailed discussion of these typological claims and more references are given in Section 4.3). While some NPs have different morphological forms for the syntactic core functions 'A' (agent in a transitive clause) and 'O' (patient), other NPs have only one form for both functions. If an NP distinguishes A and O morphologically, one of the two forms is almost always (both in terms of languages and in terms of NPs within a language) identical with the morphological form of intransitive subjects (the third core function, 'S').

Very few of the logically possible mixed case-marking patterns are actually attested in the languages of the world. If one investigates the statistical usage patterns of different transitive clause types, it turns out that those patterns that do occur are exactly those that simultaneously minimize ambiguity and speaker effort. The typology of case-marking thus appears to be a good candidate for a functional explanation.

In this paper Game Theory is used as a framework to formalize the intuitive notion of functional adaptation. It will turn out that the existing case-marking types are optimally adapted to the statistical patterns of language use in a precise game-theoretical sense.

This leaves the 'problem of linkage' (Kirby 1999) open. What are the mechanisms that enable functional pressure to shape natural languages? Working in the context of Game Theory, it is an obvious choice to employ Evolutionary Game Theory for that purpose. I demonstrate that the attested case-marking patterns are in fact those that are predicted to be evolutionarily stable.

In languages with mixed case-marking patterns, it is not entirely random which objects are marked and which aren't. Rather, case-marking applies only to a morphologically or semantically well-defined class of NPs. Take Hebrew as an example. In this language, definite objects carry an accusative morpheme while indefinite objects are unmarked.

(1) a. Ha-seret her?a ?et-ha-milxama
the-movie showed acc-the-war
'The movie showed the war.'

b. Ha-seret her?a (*?et-)milxama
the-movie showed (*acc-)war
'The movie showed a war' (from Aissen 2003)

Similar patterns are found in many languages. Bossong (1985) calls this phenomenon 'Differential Object Marking' (DOM). A common pattern is that all NPs from the top section of the definiteness hierarchy are case-marked while those from the bottom section are not.

(2) personal pronoun > proper noun > definite full NP > indefinite specific NP > nonspecific indefinite NP

Catalan, for instance, only marks personal pronouns as objects. In Pitjantjatjara (an Australian language), pronouns and proper nouns are case-marked when they are objects while other NPs aren't. Hebrew draws the line between definite and indefinite NPs and Turkish between specific and nonspecific ones.

Likewise, the criterion for using or omitting a case morpheme for objects may come from the animacy hierarchy (see for instance Comrie 1981):

(3) human > nonhuman animate > inanimate

As with the definiteness hierarchy, there are languages which mark objects only from some upper segment of this scale, like Spanish, which marks animate, but not inanimate, objects with a preposition. Finally, there are instances of DOM where case-marking is restricted to an upper segment of the product of the two scales.[2] Here an example would be Dutch where accusative marking is confined to animate pronouns.

The fact that object marking, if differential, is restricted to upper segments of the prominence scales has frequently been observed in the typological and functionalist literature. Relevant references are Aissen 2003, Bossong 1985, Comrie 1979, Comrie 1981, Dixon 1994, and Silverstein 1976. This list is far from exhaustive.

Differential case-marking also frequently occurs with subjects of transitive verbs.[3] In contradistinction to DOM, DSM ('Differential Subject Marking') means that only instances of some lower segment of the definiteness/animacy hierarchy are case-marked. The observation that the relevant scales for subjects and objects are inverses of each other is due to Silverstein (1976) (see also Comrie 1979).

DOM and DSM may co-occur within one language. This phenomenon is usually called split ergativity. This term covers both case-marking systems where the case-marking segments for subjects and for objects are complementary and systems where they overlap, i.e. where there are NP types which have both an ergative and an accusative form.

The person specification of NPs induces another hierarchy (see for instance Silverstein 1976). Simplifying somewhat, it says that the local persons (first and second) outrank third person.

(4) 1st/2nd person > third person

These patterns underlie split ergative case-marking in languages like Dyirbal where the choice between the nominative/accusative system and the ergative/absolutive system is based on person. Table 1 (taken from Aissen 1999) shows the basic case-marking pattern for Dyirbal.

[Table 1 goes about here.]

Briefly put, Dyirbal marks only local objects and third person subjects. It thus represents a combination of DOM with DSM.

These patterns of 'Differential Case Marking' (DCM) can be represented as the result of aligning two scales—the scale of grammatical functions (subject vs. object) with some scale which classifies NPs according to substantive features like definiteness, egocentricity, or animacy (as proposed in Silverstein 1976). Ranking the grammatical functions according to prominence leads to the binary scale

(5) Subj $>$ Obj

Harmonic alignment of two scales means that items which assume comparable positions in both scales are considered most harmonic. For alignment of the scale above with the definiteness hierarchy this means that pronominal subjects (+prominent/+prominent), as well as nonspecific objects (-prominent/-prominent) are maximally harmonic, while the combination of a prominent position in one scale with a nonprominent position in the other scale is disharmonic (like nonspecific subjects or pronominal objects). More precisely, harmonically aligning the hierarchy of syntactic roles with the definiteness hierarchy leads to two scales of feature combinations, one confined to subjects, and the other to objects. The subject scale is isomorphic to the definiteness hierarchy, while the ordering for objects is reversed.

(6) a. Subj/pronoun $\succ$ Subj/name $\succ$ Subj/def $\succ$ Subj/spec $\succ$ Subj/nonspec

b. Obj/nonspec $\succ$ Obj/spec $\succ$ Obj/def $\succ$ Obj/name $\succ$ Obj/pronoun

In this way DCM can be represented as a uniform phenomenon—case-marking is always restricted to lower segments of these scales.

What is interesting from a typological perspective is that there are very few attested cases of 'inverse DCM'—languages that would restrict case-marking to lower or middle segments of the above scales. The restriction to lower segments appears to be a strong universal tendency.

3 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. Which language types are functional and which aren't?

The main function of case-marking is arguably to disambiguate, i.e. to enable the hearer to identify the syntactic function of an NP. More particularly, case should uniquely identify the functions 'A' and 'O'. Now suppose that a language has two cases for NPs occurring in syntactic core roles (like nominative and accusative), and suppose case marking depends only on the syntactic role of the NP in question. Case marking thus induces a binary split of the set of roles $\{S, A, O\}$. There are four possible splits:

(7) a. $\{S, A, O\}$ vs. $\emptyset$

b. $\{S, A\}$ vs. $\{O\}$

c. $\{S, O\}$ vs. $\{A\}$

d. $\{A, O\}$ vs. $\{S\}$

The first, improper split corresponds to the absence of core case distinction, the second split to a nominative/accusative system and the third to an ergative/absolutive system. All three systems are very common among the languages of the world. The fourth system, which lumps 'A' and 'O' together and gives 'S' a different marking, is typologically virtually nonexistent.[4] As Comrie (1981) points out, there is an obvious functional motivation for this: if there are two core roles within the same clause, those are 'A' and 'P', and thus only these two need to be disambiguated. The fourth split fails to do this, and it introduces an entirely superfluous distinction between roles that can never be confused anyway. The first system can be argued to be dysfunctional as well, but languages without core case distinctions have other means to disambiguate syntactic roles, e.g. head marking and word order.

A tripartite split, which gives a different case for each core function, induces a superfluous distinction between 'S' and the two transitive functions. Consequently, such systems are equally rare.

There is thus good functional motivation for the only basic case-marking systems being the nominative/accusative system and the ergative/absolutive system. [5] If there are two competing grammatical forms within a paradigm, one is usually more marked than the other. Taking this for granted, we still have four basic case-marking systems to consider:

(8) a. unmarked nominative vs. marked accusative

b. unmarked nominative vs. marked ergative

c. marked nominative vs. unmarked accusative

d. marked nominative vs. unmarked ergative

At least since Zipf (1949) it is a standard assumption that speakers strive to minimize effort in language production. [6] This means that, all else being equal, speakers should prefer the unmarked over the marked form. From this it follows directly that only the first two patterns are functional. In the systems using a marked nominative, the speaker has to use a marked form to express 'S', i.e. an entirely unambiguous role. It would be more economical for the speaker to use the unmarked form there instead. These functional considerations also square nicely with the typological facts. As Dixon (1994:63) writes: 'S is normally unmarked, since there is no other NP in an intransitive clause from which it must be distinguished; it then falls together with the unmarked transitive function.' There are exceptions to this rule, i.e. languages with a marked nominative, but they are rare (see the discussion in Dixon 1994:63–67). Further support comes from Greenberg's (1963) Universal 38: 'Where there is a case system, the only case which ever has only zero allomorphs is the one which includes among its meanings that of the subject of the intransitive verb.'

We can thus assume that the only fully functional case-marking systems are accusative systems and ergative systems with an unmarked nominative. If we take the option of differential case-marking into account, further functionality considerations come into play. We can assume without loss of generality that the hearer always interprets ergative as A and accusative as O (provided the case-marking within a clause is consistent), so ambiguity can safely be avoided if at least one NP per clause is case-marked. Differential case marking means that the presence or absence of case-marking on an NP depends on a certain feature of that NP. Suppose NPs can be characterized as 'prominent' or 'nonprominent' according to some notion of prominence (i.e. some cut-off point at one of the prominence scales mentioned above). For the sake of concreteness, let us assume that case-marking patterns may be different for pronouns and full NPs. Following Silverstein, I take pronouns to be more prominent than full NPs, and thus denote the pronoun class by 'p' (for 'prominent') and the full NPs by 'n' (nonprominent). The following mappings from syntactic functions to case morphemes then avoid ambiguity. In what follows, 'e' abbreviates 'ergative', 'a' accusative, and 'z' 'zero', i.e. the nominative form.

(9)
a. $A \to e, O \to a$
b. $A/p \to e, A/n \to z, O \to a$
c. $A/p \to z, A/n \to e, O \to a$
d. $A \to z, O \to a$
e. $A \to e, O/p \to a, O/n \to z$
f. $A \to e, O/p \to z, O/n \to a$
g. $A \to e, O \to z$

For the sake of brevity, from now on such patterns are denoted as a quadruple of case forms, in the order $A/p, A/n, O/p, O/n$. The above patterns thus become $eeaa$, $ezaa$, $zeaa$, $zzaa$, $eeaz$, $eeza$, $eezz$.

All other patterns will have $z - z$ clause types, i.e. certain clause types where both arguments are zero marked. This need not lead to ambiguity if one of the two unmarked arguments is prominent and the other isn't. Then the hearer may employ a default rule to the effect that in such a case the more prominent NP is A (or vice versa). With 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 syntactic 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. So the mentioned nine strategies are the only ones that guarantee 100% disambiguation.

Let us assume that unambiguous encoding is the main priority of the speaker, but he has the secondary priority to use as few case morphemes as possible. Of the mentioned case marking patterns, five use redundant case-marking. The only ones that distinguish between A and O and use minimal case morphology are: $eezz, zzaa, zeaz, ezza$.

Which of these patterns minimizes the average number of case morphemes per clause depends on the frequencies of clause types. We only have to consider four clause types – both A and O may be p or n. The figures in Table 2 are extracted from Geoffrey Sampson's CHRISTINE corpus of spoken English.[7]

[Table 2 goes about here.]

(I refer to the four cells of such a table with pairs pp, pn, np, and nn, where the first element gives the specification of A and the second of O.) Under the assumption that this is a representative sample, the expected numbers of case morphemes per clause for the four strategies are (where $N = pp + pn + np + nn$ is the total number of transitive clauses in the corpus):

(10)
$$\begin{array}{lll} eezz: & (pp+pn+np+nn)/N & = 1.0 \\ zzaa: & (pp+pn+np+nn)/N & = 1.0 \\ zeaz: & (pp+2np+nn)/N & \approx 0.30 \\ ezza: & (pp+2pn+nn)/N & \approx 1.70 \end{array}$$

Obviously the harmonically aligned split ergative pattern $zeaz$ turns out to be optimal. This fact does not depend on the particular numbers but just on the inequality

(11) $pn > np$

That is, there are more instances of clauses with prominent agent and nonprominent object than clauses with nonprominent agent and prominent object. This inequality holds for all splits along the definiteness scale or the person scale. In Table 3 you find the figures for the other split points (from CHRISTINE).

[Table 3 goes about here.]

I did the same tests with a corpus of spoken Swedish.[8] The annotation included animacy. There the above inequality holds as well – the precise numbers are given in Table 4.

[Table 4 goes about here.]

It is instructive to look at the figures for definiteness in the same corpus of spoken Swedish, which can be found in Table 5.

[Table 5 goes about here.]

Here the definite objects (1830) outnumber the indefinite objects (1321). So at first glance one might assume that it would be more economical to restrict accusative marking to indefinites than to definites. However, the above considerations show that a split ergative pattern with ergative marking of indefinite subjects and accusative marking of definite objects would be optimal for such a frequency distribution as well because $pn > np$.

4 Game Theory

4.1 The utility of case-marking The notion of optimality used in the previous section was defined in a rather ad hoc fashion. Game Theory is well-suited to make things more precise. Game Theory (GT henceforth) was developed as a model of rational decision making in interaction. Every participant in a game receives a certain payoff (which is defined numerically and may be negative). Crucially, the payoff of each player depends both on his own action and on the actions of the other players.

In the context of linguistics, a game can be identified with a single utterance situation. Speaker and hearer are the players. Their actions are the production and interpretation of an utterance respectively, and their payoff preferences correspond to speaker economy and hearer economy.

To makes things 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 h 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, some random device, usually called nature by game theorists, presents the speaker with a meaning m, and the speaker only has to choose how to express m.[9]

Communication is successful if the hearer recovers the intended meaning from the observed form. It is measured by the δ-function:[10]

(12)
$$\delta_m(s,h) = \begin{cases} 1 & \text{iff} \quad h(s(m)) = m \\ 0 & \text{otherwise} \end{cases}$$

In words, the δ-value is 1 if the hearer recognizes the meaning that is intended by the speaker, and 0 otherwise.

Forms differ with respect to their complexity. I take it that this complexity can be measured numerically, i.e. cost is a function from F to the nonnegative real numbers.

Speakers have two possibly conflicting interests: they want 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:[11]

(13)
$$u_s(m,s,h) \quad = \quad \delta_m(s,h) - k \times \underline{\text{cost}}(s(m))$$

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:

(14) $$u_h(m, s, h) \quad = \quad \delta_m(s, h)$$

One might argue that complexity in interpretation also incurs processing costs for the hearer (which need not coincide with the costs for the speaker), and that these costs should be reflected in the hearer's utility function. There is a fundamental asymmetry between speaker and hearer though. The speaker has a choice between using more or less complex forms. The hearer has no such choice; he has to process the form that he gets from the speaker. Neither does he have the choice between more or less complex interpretations; his goal is to figure out what the speaker is trying to communicate, no matter how complex this meaning may be. What is important in game theory is not the absolute utility but the utility differences between strategies. Since no choice on the hearer's side can possibly impact his processing load, hearer complexity can safely be left out of consideration.

Nature presents meanings to the speaker according to a certain probability distribution P that determines how likely a certain meaning is to be expressed. It represents cognitive and communicative tendencies, not peculiarities of a certain language. The average utilities of speaker and hearer in a game can thus be given as

(15) $$u_s(s, h) \quad = \quad \sum_m P(m) \times (\delta_m(s, h) - k \times \underline{\text{cost}}(s(m)))$$

(16) $$u_h(s, h) \quad = \quad \sum_m P(m) \times \delta_m(s, h)$$

Let us see how case-marking patterns can be described by game-theoretic notions. We are only concerned with elementary transitive clauses. [12] So we are dealing with two NPs. One is

A and the other O, and both may be either p or n according to some split point on the definiteness or animacy scale. I am not concerned with the effect of word order or head marking on argument linking in this article. So we consider only those clauses where the assignment of NPs to roles is ambiguous unless case-marking is taken into account. How many clauses have this property depends on a variety of factors, most notably on the word order regularities of the language in question. In English there are very few such constructions. One example would be comparative deletion, as in

(17) John saw Bill earlier than Joe.

Here the interpretation of Joe as subject or object is not grammatically disambiguated. This kind of ambiguity is more pervasive in languages with freer word order. German has a strong preference to place A before O. In relative clauses, this does not help for disambiguation because the verb is clause-final and the relative pronoun is clause-initial in any case. The following example is thus fully ambiguous. (The nouns and pronouns here are all feminine, and German does not differentiate between nominative and accusative in feminine gender. Therefore here even case does not help to disambiguate.)

(18) die Frau, die Maria gesehen hat
the woman, which Maria seen has
'the woman which Maria saw/which saw Maria'

In languages with very free word order like Latin, this kind of ambiguity (ignoring case-marking) is the norm rather than the exception even in regular main clauses.
So we consider only clauses here that are not disambiguated by word order or head marking. Therefore I take it that nature chooses the word orders $A - O$ and $O - A$ with a 50 % probability each,[13] and that this choice is stochastically independent from the specifications of the NPs with 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:[14]

(19) $A/p - O/p$

$A/p - O/n$

$A/n - O/p$

$A/n - O/n$

$O/p - A/p$

$O/p - A/n$

$O/n - A/p$

$O/n - A/n$

P is a probability distribution over these eight meanings. The stochastic independence of word order from all other factors entails that for all $i, j \in \{p, n\}$:

(20) $$P(A/i - O/j) \;=\; P(O/j - A/i)$$

Now reconsider the corpus data discussed in the previous section. Let us assume that the relative frequencies of pp, pn, np and nn in a given corpus (for a given instantiation of p and n) are accurate estimates of the underlying probabilities given by nature. Let $N = pp + pn + np + nn$ be the total number of transitive clauses in a corpus. Then we have for all $i, j \in \{p, n\}$:

$$\frac{ij}{N} = P(A/i - O/j) + P(O/j - A/i) = 2P(A/i - O/j)$$

It is plausible to assume that the prominence of an NP is always unambiguously encoded in its form. Except for the contrast between definites and indefinites, it is hard to imagine a language where this is not the case. As for definiteness, I assume for simplicity that it is always either morphologically or syntactically marked as well.[15] 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:

(21) $$\begin{aligned} \underline{\text{cost}}(e/i - e/j) = \underline{\text{cost}}(e/i - a/j) = \underline{\text{cost}}(a/i - e/j) = \underline{\text{cost}}(a/i - a/j) &= 2 \\ \underline{\text{cost}}(e/i - z/j) = \underline{\text{cost}}(a/i - z/j) = \underline{\text{cost}}(z/i - e/j) = \underline{\text{cost}}(z/i - a/j) &= 1 \\ \underline{\text{cost}}(z/i - z/j) &= 0 \end{aligned}$$

That is, case-marking both agent and object is most costly, and zero marking both core arguments is most economical. Of course a clause contains other sources of complexity than just case morphology, but in the present context only the difference between case variants of the same clause type are relevant, and we can thus neglect all other costs.

For each of the eight meanings given in (19), the speaker can in principle choose out of the nine different forms given in (21). (The values of i and j are determined by the meaning.) This would give us a total of $9^8 = 43,046,721$ different speaker strategies. However, I limit attention to just a small subset of simple strategies. First, word order effects are kept out of consideration. So if meaning $m2$ results from exchanging the order of the NPs in meaning $m1$ while keeping their prominence values and syntactic functions unchanged, then $s(m2)$ should be the mirror image of $s(m1)$. 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.[16] There are just 81 strategies with this property. Among these strategies, I 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:

(22) $eeaa$, $eeaz$, $eeza$, $eezz$, $ezaa$, $ezaz$, $ezza$, $ezzz$, $zeaa$, $zeaz$, $zeza$, $zezz$, $zzaa$, $zzaz$, $zzza$, and $zzzz$

The range of these speaker strategies only contains consistent case-markings; neither double ergative nor double accusative ever occur. Hence, of the 36 possible forms mentioned above, we have to consider only 28, since we can exclude $e/i - e/j$ and $a/i - a/j$ for all possible values of $i, j \in \{p, n\}$.

A hearer strategy is a mapping from forms to meanings. For each form there are two meanings that are consistent with the prominence values of the NPs, $A - O$ and $O - A$ word order respectively. With 28 forms this leads to $2^{28} = 268,435,456$ different hearer strategies. Here too we can safely exclude the vast majority from the outset. 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 use the designation ‘faithful’ for the hearer strategies that interpret ergative as A and accusative as O.

Faithfulness leaves the interpretation of clauses with the case-marking pattern $z - z$ open. The only clues for the hearer in these cases are the prominence values of the arguments, and word order. There are four such clause types (depending on the prominence features of the two NPs), each of which may receive two possible interpretations. Therefore we have 16 faithful hearer strategies. 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 $n/z - p/z$. These strategies are denoted by AO, pA, pO and OA. AO is the strategy that assumes the default word order $A - O$, and likewise for OA. The strategy pA assumes the default that if in doubt, the more prominent argument is the agent, and pO assumes the more prominent argument to be the object.

4.2 Rationalistic Game Theory The classical paradigm of GT, as conceived by von Neumann, Morgenstern, Nash and other theoreticians in the mid-twentieth century, is rationalistic.[17] The central question is: given a game, i.e. a strategy set for each player and a utility function, which strategy should a perfectly rational player choose?

A very simple and powerful concept for solving this problem is the notion of strict strategy domination. A strategy s_1 strictly dominates strategy s_2 if a rational player always prefers s_1 over s_2. (A rational player is a player who is logically omniscient and only wants to maximize his utility.) For two speaker strategies, this means formally

$$s_1 > s_2 \Leftrightarrow \forall h : u_s(s_1, h) > u_s(s_2, h)$$

This means that the speaker strategy s_1 does better than s_2, no matter which strategy the hearer uses. The notion of strict domination between hearer strategies is entirely analogous. Strategy h_1 strictly dominates h_2 if the former outperforms the latter against each speaker strategy; formally

$$h_1 > h_2 \Leftrightarrow \forall s : u_h(s, h_1) > u_h(s, h_2)$$

A rational player will never play a strictly dominated strategy.

We can now identify several strictly dominated speaker strategies. To start with, $eeaa$, $eeza$, $eeaz$, $ezaa$ and $zeaa$ are strictly dominated. Take $eeza$ and compare it with $eezz$. Both strategies exclude ambiguity, so the δ-function will always be 1. However, $eeza$ uses two case morphemes if O is nonprominent, while $eezz$ always uses exactly one case morpheme per clause. So the expected costs of $eezz$ are lower and thus its average utility is higher. In other words, $eezz$ strictly dominates $eeza$. The same kind of reasoning applies to the other four mentioned strategies.

The 'inverse split ergative' strategy $ezza$ is also strictly dominated. It uses two morphemes for the meaning $A/p - O/n$, zero morphemes for $A/n - O/p$, and one morpheme for the other meanings. So its expected costs are

$$\frac{pp}{N} + 2 \times \frac{pn}{N} + \frac{nn}{N} = 1 + \frac{pn - np}{N}$$

Since $pn > np$, the expected costs of $ezza$ exceed 1. The pure ergative strategy $eezz$, for instance, has expected costs of exactly 1. The δ-function for $eezz$ is always 1, while for $ezza$ it may be lower than 1, depending on the hearer strategy. Hence $eezz$ strictly dominates $ezza$, and we may eliminate the latter.

The utility functions of the remaining game—comprising ten speaker strategies and four hearer strategies—are given in Table 6. The rows represent speaker strategies and the columns hearer strategies. The cells contain the hearer utility of the corresponding strategy pair. The last column holds the expected costs of each speaker strategy, multiplied with the coefficient k. The speaker utility for each cell can be obtained by subtracting the last column from the corresponding hearer utility.

[Table 6 goes about here.]

What is the linguistic interpretation of the parameter k? Loosely speaking, a high value of k means 'costly morphology' and vice versa. It is difficult to quantify the relative importance of clarity versus brevity. In certain registers people use very elaborate formal means to avoid ambiguity, so there is no obvious upper limit of k. k is correlated with an obvious linguistic parameter though. If a speaker strategy (i.e. a production grammar) uses accusative marking of pronouns, say ($zzaz$), it will use it in all transitive clauses,[18] no matter whether the case-marking is necessary to disambiguate or not. Case marking will thus be more useful in languages with free word order (many potentially ambiguous clauses) than in languages with strict word order. To put it differently, the relative costs of case-marking (relative to its benefit for disambiguation) is high in languages with strict word order and low in languages in free word order. The value of k can thus be seen as being correlated with the degree of word order strictness. Languages with free word order have a low value for k and vice versa. This of course only holds everything else being equal. Despite the strict word order of English, for instance, the differential complexity of the nominative and accusative forms of the pronouns (like he vs. him) is so small—(unmarked) open syllable versus (marked) closed syllable—that the value of k for English is still not very high.

For concreteness, let us fill in some numerical values. In Table 7, the speaker utilities are given that result if the distributions of pronouns ($= p$) vs. full NPs ($= n$) in the CHRISTINE corpus is taken to be representative, and the cost coefficient k is set to 0.1. This would correspond to a language with very free word order. To obtain the utility matrix for the hearer, one would have to subtract the costs of each strategy pair from each cell. However, the hearer can only choose between cells on the same row, and the costs are identical within one row. Hence the relative utilities of the different strategies for the hearer remain the same if we identify the hearer utility with the speaker utility, and this is followed from now on.

19

[Table 7 goes about here.]

Not surprisingly, the strategy pair $zeaz/pA$, i.e. split ergativity, turns out to yield the maximal payoff, namely 0.97. But what are the predictions of rationalistic GT?

The most important concept of rationalistic GT is the notion of a Nash Equilibrium ('NE' henceforth). This is a pair of strategies with the property that each of the two is a best response to the other. Applied to our linguistic games, this means formally:

Definition 1 (Nash Equilibrium). A pair of strategies (s, h) is a Nash Equilibrium iff

$$\forall s' : u_s(s,h) \geq u_s(s',h)$$

and

$$\forall h' : u_h(s,h) \geq u_h(s,h')$$

One can expect that rational players will always play NE strategies because in a NE, neither player has an incentive to change his strategy unilaterally.

In a payoff table as Table 7, a cell represents an NE iff its content is highest both in its row and its column. There are seven NEs in Table 7:

(23) a. $zeaz/pA$

b. $eezz/AO$, $eezz/pO$, $eezz/OA$

c. $zzaa/AO$, $zzaa/pO$, $zzaa/OA$

Some tedious but elementary calculations show that these seven strategy pairs are always NE provided $pp > nn$ and $k < \frac{nn}{2(np+nn)}$, or $nn > pp$ and $k < \frac{pp}{2(pp+pn)}$

It is not difficult to see that these NEs are in fact intuitively reasonable. Suppose the hearer uses the correct default that in a clause containing a prominent and a nonprominent argument, the prominent argument is A and the nonprominent O (i.e. the strategy pA). In this case, the speaker can safely omit case-marking whenever the meaning to be expressed matches this default. Case marking is only necessary if the default is not applicable or would yield the wrong result, and $zeaz$ is the only way to implement this. On the other hand, if the speaker uses $zeaz$, the hearer will always get the correct meaning if he uses pA, and thus can only lose by changing his strategy.

Suppose, on the other hand, the hearer uses the opposite default (pO) or some mixed default (AO or OA). Then the speaker has to disambiguate each clause with at least one case morpheme, and the most economical way to do so is to use exactly one morpheme per clause. $eezz$ and $zzaa$ are the only strategies that do exactly that. If the speaker uses one of those strategies, the hearer will always get the right meaning and has thus no incentive to change his strategy.

As mentioned above, this configuration of NEs only holds if k is comparatively small, i.e. the word order is free. Let us explore what happens if k is higher.

In this case, the proportion of prominent As on the one hand and nonprominent Os on the other hand makes a difference. For the distinction between local person vs. third person in CHRISTINE, $nn > pp$, while all other configurations lead to $pp > nn$. This may be a peculiar feature of the particular corpus. However, it is rather obvious that the proportion $pp : nn$ will be low if the split point is high in the hierarchy and vice versa. To illustrate this point, suppose we calculate these figures for two split points—local persons vs. third person, and definite vs. indefinite NPs, say—and the same corpus. For the first split, pp_1 would be the number of transitive clauses where both subject and object are first or second person pronouns, while nn_1 is the number of clauses where both arguments are third person. pp_2, on the other hand, is the number of clauses where both arguments are definite, and nn_2 the number of clauses comprising two indefinite NPs. Since each local person NP is definite (and thus each indefinite NP third person), it necessarily holds that $pp_1 < pp_2$, and $nn_1 > nn_2$. Hence, if $pp_1 > nn_1$, then $pp_2 > nn_2$. By the same kind of reasoning it is clear that whenever the first split point is higher on the hierarchy than the second, then $pp_1 : nn_2 < pp_2 : nn_2$. So it is reasonable to assume that for high split points, $nn > pp$ is more likely while for low split points, the opposite is true.

Sticking to the figures from CHRISTINE and pronoun/full NP, it holds that $pp > nn$. The critical quantity for k is ≈ 0.412. Table 8 contains the numbers for $k = 0.45$.

[Table 8 goes about here.]

There are six NEs in this configuration:

(24) a. $zzaz/pA$

b. $eezz/AO$, $eezz/OA$

c. $zzaa/AO$, $zzaa/OA$

d. $ezzz/pO$

The first one is the typologically very common DOM configuration where accusative marking is confined to prominent Os and no ergative morpheme is used. Besides, both pure ergative and pure accusative marking is an NE, but only in combination with a hearer strategy that does not employ prominence for interpretation. Finally, there is another NE, which is very rare among the languages of the world: an inverse form of DSM, i.e. prominent As are ergative marked while Os are always unmarked.[20]

This configuration of NEs obtains if $pp > nn$ and $\frac{nn}{2(np+nn)} < k < 0.5$. If k grows even larger, different NEs emerge. If $0.5 < k < \frac{pp+2np}{2(pp+np)}$, we have two different NEs again. (The figures for $k = 0.53$ are given in Table 9.)

[Table 9 goes about here.]

Now DSM ($zezz/pA$) has a higher average utility than DOM ($zzaz/pA$), and it emerges as NE. Likewise, $zzza$ now outranks $ezzz$ if the hearer strategy is pO, and thus becomes an NE. These are the only NEs in this game.

If the costs are still higher, the speaker has little incentive to use case-marking. Consider the situation with $\frac{pp+2np}{2(pp+np)} < k < \frac{2pn+nn}{2(pn+nn)}$. If the hearer strategy is pA, the most economical response for the speaker is $zzzz$, i.e. no case-marking at all. As response to pO, inverse DOM ($zzza/pO$) is still optimal. Table 10 gives the values for $k = 0.7$.

[Table 10 goes about here.]

Finally, if k grows extremely high ($k > \frac{2pn+nn}{2(pn+nn)} \approx 0.95$), it never pays for the speaker to use case morphology. The speaker strategy $zzzz$ strictly dominates all other strategies, and the best response to $zzzz$ is pA, so there is only one NE. Table 11 gives the values for $k = 1$.

[Table 11 goes about here.]

It is perhaps important to stress that these configurations of Nash Equilibria do not depend on the particular numbers used here. All that matter is that $pn > np$ and $pp > nn$. As the distribution of local person vs. third person NPs in CHRISTINE shows, the latter inequality need not hold. If $nn > pp$, the structure of NEs is essentially a mirror image of the scenarios discussed so far. For very small values of k, we have the same seven NEs as with $pp > nn$, namely $zeaz/pA$, and all combinations of a pure strategy ($eezz$ or $zzaa$) with any hearer strategy $\neq pA$. If k grows larger ($\frac{pp}{2(pp-np)} < k < 0.5$), both DSM ($zezz/pA$) and inverse DOM ($zzza/pO$) are NEs, next to $eezz/AO$, $eezz/OA$, $zzaa/AO$ and $zzaa/OA$. For $0.5 < k < \frac{2np+nn}{2(np+nn)}$ we find the NEs DOM ($zzaz/pA$) and inverse DSM ($ezzz/pO$). If $\frac{2np+nn}{2(np+nn)} < k < \frac{pp+2pn)}{2(pp+pn)}$, uniform zero marking ($zzzz/pA$) and inverse DSM are each in equilibrium, and for very large values of k ($k > \frac{pp+2pn)}{2(pp+pn)}$), again only $zzzz/pA$ is NE.[21]

4.3 Taking stock Let us see how this relates to the findings of typological research. 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. According to Blake (2001), there is apparently only one language with a full-blown tripartite system, i.e. with the strategy $eeaa$, namely the Australian language

Wangkumara (see Breen 1976, cited after Blake 2001). Inverse split ergative systems—$ezza$ in my system—are also very rare. Plank (2001) mentions the Iranian languages Parachi and Yazguljami as examples though. 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.

Finally, there are no languages which would have a tripartite paradigm for all and only the prominent or all and only the nonprominent NPs. Hence $zeza$ and $ezaz$ are correctly excluded. Conversely, we expect to find instances of languages with an NE pattern. This is certainly the case for DOM $zzaz$ (like English or Hebrew), DSM $zezz$—for instance Yup'ik Eskimo, (Reed et al. 1977, cited after Blake 2001) or the Circassian languages Adyghe and Kabardian Kumakhov et al. 1996—split ergative $zeaz$ (like Dyirbal, see Dixon 1972), and zero marking like the Bantu languages.

So far the concept of a Nash Equilibrium proves fairly successful in identifying possible case-marking systems. Strategies that do not participate in an NE are in fact nonexistent or very rare. However, the concept is still too inclusive. I know of only one language of the type $zzza$, namely Nganasan (see Dixon 1994, p. 90), and no undisputed instance of $ezzz$ (but see footnote 20). The pure accusative system—$zzaa$—does exist (Hungarian would be an example), but it too is very rare. Most accusative languages have DOM, and most ergative languages DSM. As Blake (2001) puts it:

> 'Pure accusative systems of marking noun phrases where the marking of the object is always distinct from the marking of the subject are rare'. (p. 119)[22]

and

> 'In languages with ergative case-marking on nouns it is true more often

> than not that the ergative marking is lacking from first- and second-person pronouns and sometimes from third.' (p. 122)

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.

4.4 Evolutionary Game Theory Evolutionary Game Theory (EGT) was developed by theoretical biologists, especially John Maynard Smith (cf. Maynard Smith 1982) as a formalization of the neo-Darwinian concept of evolution via natural selection. It builds on the insight that many interactions between living beings can be considered to be games in the sense of GT—every participant has something to win or to lose in the interaction, and the payoff of each participant can depend on the action of all other participants. In the context of evolutionary biology, the payoff is an increase in fitness, where fitness is basically the expected number of offspring. According to the neo-Darwinian view on evolution, natural selection operates on genetically determined—and thus heritable—traits of individuals. If the behavior of interactors in a game-like situation is genetically determined, the strategies can be identified with gene configurations.

In the EGT setting, we are dealing not just with one game and its participants, but with large populations of potential players. Each player is programmed for a certain strategy, and the members of the population play against each other very often under total random pairings. The payoffs of each encounter are accumulated as fitness, and the average number of offspring per individual is proportional to its accumulated fitness, while the birth rate and death rate are constant. Parents pass on their strategy to their offspring basically unchanged.[23] If a certain strategy yields on average a payoff that is higher than the population average, its replication rate will be higher than average and its proportion within the overall population increases, while strategies with a less-than-average expected payoff decrease in frequency. A strategy mix is stable under replication if the relative proportions of the different strategies within the population do not change under replication.

Occasionally replication is unfaithful though, and an offspring is programmed for a different strategy than its parent. If the mutant has a higher expected payoff (in games against members of the incumbent population) than the average of the incumbent population itself, the mutation will spread and possibly drive the incumbent strategies to extinction. For this to happen, the initial number of mutants may be arbitrarily small.[24] Conversely, if the mutant does worse than the average incumbent, it will be wiped out and the incumbent strategy mix prevails.
A strategy mix is evolutionarily stable if it is resistant against the invasion of small proportions of mutant strategies. In other words, an evolutionarily stable strategy mix has an invasion barrier. If the amount of mutant strategies is lower than this barrier, the incumbent strategy mix prevails, while invasions of higher numbers of mutants might still be successful.
In the metaphor used here, every player is programmed for a certain strategy, but a population can be mixed and comprise several strategies. Instead we may assume that all individuals are identically programmed, but this program is nondeterministic and plays different strategies according to some probability distribution (which corresponds to the relative frequencies of the pure strategies in the first conceptualization). Game theorists call such nondeterministic strategies mixed strategies. For the purposes of the evolutionary dynamics of populations, the two models are equivalent. It is standard in EGT to talk of an evolutionarily stable strategy, where a strategy can be mixed, instead of an evolutionarily stable strategy mix. This terminology is followed henceforth.
The notion of an evolutionarily stable strategy can be generalized to sets of strategies. A set of strategies A is stationary if a population where all individuals play a strategy from A will never leave A unless mutations occur. A set of strategies is evolutionarily stable if it is resistant against small amounts of non-A mutants. Especially interesting are minimal evolutionarily stable sets, i.e. evolutionarily stable sets which have no evolutionarily stable proper subsets. If the level of mutation is sufficiently small, each population will approach such a minimal evolutionarily stable set. In the games discussed in this article, all minimal evolutionarily stable sets consist of just one speaker- and one hearer-strategy, so that we can restrict attention to the simpler notion of evolutionarily stable strategy pairs.

How can this model be applied to linguistics? If the strategies in the EGT sense are identified with grammars (as done in the previous subsection), games should be identified with utterance situations. However, grammars are not transmitted via genetic but via cultural inheritance (within the boundary of innate grammatical knowledge, which is not the topic of this article). Therefore, imitation dynamics[25] are more appropriate here than the replicator dynamics that are used in applications of EGT to theoretical biology. According to imitation dynamics, players are not mortal and have no offspring. However, every so often, a player is offered the opportunity to pick out some other player x and to replace his own strategy by the strategy of x. 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 previous model, successful (in terms of utility) strategies will tend to spread while unsuccessful strategies die out. Those (possibly mixed) strategies that are evolutionarily stable under the imitation dynamics are exactly the same that are stable under the replicator dynamics.

Imitation dynamics are compatible with several linguistic interpretations. First, one might assume that each speaker follows a certain deterministic strategy (something like a certain parameter configuration in a Principles-and-Parameters framework in the sense of Chomsky 1981), but speakers/hearers are allowed to switch parameters. Under this interpretation, imitation dynamics amount to learning via imitation in a rather literal sense.

Second, one might assume that language users use a mixed strategy—either in the form of a probabilistic grammar (as frequently assumed in computational linguistics and socio-linguistics) or in the sense of probabilistic diglossia (as is sometimes assumed in generative approaches to language change, see for instance Kroch 2000). Then imitation dynamics amount both to learning by imitation and learning by experience. Of course, under this interpretation the players are mortal and may have offspring. However, this interpretation of the evolutionary dynamics of language is primarily usage-based, and the effect of the biological replacement of individuals (and concomitant language acquisition) may legitimately be neglected as a first approximation.

Third, imitation dynamics are compatible with the Iterated Learning Model of language evolution (Kirby 1999, Nowak et al. 2002), according to which language evolution is driven by first language acquisition. Under this interpretation one has to assume that death rate and birth rate are approximately constant. The death of an individual and the language acquisition by the infant that replaces it can be seen as an imitation process. For this interpretation one has to assume that grammars with a high expected utility are more likely to be acquired by infants than less successful grammars.

The second interpretation (speakers use stochastic grammars and learn by imitation and experience) strikes me as most plausible. Here the evolutionary dynamics essentially amount to the adjustment of grammatical preferences of adult speakers. Possible sources of variation that are relevant in the present connection include phonological reduction under redundancy, and periphrasis to avoid ambiguity (see also the more thorough discussion of this point in section 5), as well as socio-linguistic factors like language contact.

However, the formal results to be reported below are compatible with the other interpretations as well.

Under each interpretation, we expect that most natural language grammars are evolutionarily stable because unstable grammars do not persist. It is perhaps important to stress once more that the main explanatory value of imitation dynamics, and of EGT in the present context in general, is not to explain or predict language change. Grammars are basically predicted to be evolutionarily stable. We expect diachronic change to occur either if a mutation exceeds the invasion barrier of an evolutionarily stable strategy, or else if the fitness landscape itself changes, i.e. if either the range of available strategies or the utility function changes. EGT can be used to study the consequences of such incidents, but not to explain them as such. So the entire approach laid out here is relevant for typology rather than for historical linguistics. This claim may perhaps seem paradoxical at a first glance—the linguistic instantiation of the evolutionary dynamics is language change of some sort, so historical linguistics must be involved somehow, one should think. My point here is that the model says nothing about the precise mechanisms and trajectories of language change. All that matters here are the likelihood of the state of a language to be the source or the target of change. Let me use an analogy from biology to bring this point home. As mentioned above, a sex ratio of 1:1 is the only evolutionarily stable state of a population from a sexually reproducing species. Nonetheless, slight male or female biases constantly occur, for a variety of reasons. However, if a population consists of more female than male individuals (in reproductive age), males have on average a higher chance to mate successfully than females, and thus males have a higher expected number of offspring. Consequently, a heritable disposition to spawn male offspring will spread more quickly in the population than a disposition to spawn female offspring, until the 1:1 balance is restored. By a symmetrical argument, the same holds for a population with a male bias. The precise biochemical and ethological mechanisms that are the proximate causes of these processes may differ from species to species. The considerations that only a 1:1 ratio is evolutionarily stable does not explain these processes, but it does not rely on a precise understanding of them either. All that matters is that there are heritable dispositions for spawning offspring of a particular sex.

Let us return to the evolutionary dynamics of case-marking systems. The notion of evolutionary stability only makes sense if there are diachronic processes that lead to the emergence or to the loss of case-marking for certain NP types. Historical linguistics has established ample evidence for that—more about this later. Considerations of evolutionary stability cannot be used to make predictions about the precise diachronic processes. On the other hand, the notion of stability can be established independently from the precise nature of these processes.

The Game of Case that was introduced in the last subsection is an asymmetric game. Every player is either speaker or hearer, and speakers and hearers have different strategy sets at their disposal. In a population dynamic setting, this means that we are dealing with two separate populations—the speakers and the hearers.[26] So rather than with evolutionarily stable strategies, we have to deal with evolutionarily stable strategy pairs here. In multi-population dynamics, evolutionary stability can be characterized quite easily in rationalistic terms.[27] Briefly put, a strategy pair is evolutionarily stable iff it is a Strict Nash Equilibrium (SNE henceforth). For an NE to be strict, each of its components must be the unique best reply to the other component.

Definition 2 (Strict Nash Equilibrium). A pair of strategies (s, h) is a Strict Nash Equilibrium iff

$$\forall s'(s' \neq s \rightarrow u_s(s,h) > u_s(s',h))$$

and

$$\forall h'(h' \neq h \rightarrow u_h(s,h) > u_h(s,h'))$$

Theorem 1 (Selten 1980). (s, h) is evolutionarily stable if and only if it is a Strict Nash Equilibrium.

Let us apply the analytical tools of EGT to the different instantiations of the Game of Case. I start with the first version where k is very small. The utility function is illustrated in Table 7.[28] As pointed out in the previous section, it has seven NEs. However, only one of them is strict, namely the split ergativity pattern:

- $zeaz/pA$

The other six NEs are not evolutionarily stable. To see why this is so, consider the nonstrict NE $zzaa/AO$. (The following considerations apply to the other five nonstrict equilibria as well.) For simplicity, I consider only populations where the speakers use either $zeaz$ or $zzaa$, and the hearers either pA or AO. Suppose the speaker population uses almost exclusively $zzaa$, and the hearers almost always AO. Then the speaker strategy $zzaa$ does on average better against the hearers than $zeaz$ (slightly less than 0.90 against slightly more than 0.61). Hence we expect that $zzaa$ is imitated more often than $zeaz$, and the latter is wiped out in the long run. Why is this state not evolutionarily stable?

The dynamics of the game with just the two speaker strategies and two hearer strategies mentioned are as indicated in Figure 1, where the vector field and some sample trajectories are depicted. Each point in the square represents a certain mix of speaker strategies and hearer strategies within the population. The position along the x-axis indicates the proportion of $zeaz$-speakers (from 0% at the left edge to 100% at the right edge) and the position along the y-axis the proportion of pA-hearers (bottom corresponds to 0% and top to 100%).

[Figure 1 goes about here.]

There are in fact two regions where trajectories end—the SNE $zeaz/pA$ (the upper right corner, indicated by a black square), but also the lower part of the left boundary of the graphics (indicated by a bold vertical line). This corresponds to the combination of the speaker strategy $zzaa$ with any combination of the two hearer strategies, as long as the proportion of pA is below a certain critical value.

However, these dynamics do not take the effect of possible mutations into account. Suppose a pure $zzaa/AO$ population combination is invaded by a small amount of $zeaz/pA$ mutants. Then the $zeaz$-speaker will swiftly be wiped out because they do much worse against the hearer population average than the incumbent $zzaa$-strategy. However, the hearer mutants pA do as well against the incumbent speakers as the incumbent hearers (the utility is 0.90 for both), and they do much better against the mutant speakers as the incumbent hearers (0.97 against 0.61). So as a consequence of the invasion, the mutant speaker will be wiped out but the mutant hearers survive and even spread. If such an invasion occurs again, the pA-strategy will further extend its share of the population. After finitely many repetitions of this mutant invasions, the proportion of pA-hearers will exceed the critical threshold and the entire population is pushed into the basin of attraction of $zeaz/pA$.

If one assumes an arbitrarily small but constant supply of mutants, the overall dynamics changes its character drastically. There is just one attractor, namely the SNE $zeaz/pA$. The dynamics are sketched in Figure 2.

[Figure 2 goes about here.]

If $pp > nn$ and $\frac{nn}{2(np+nn)} < k < 0.5$, we get a configuration as in Table 8. Of the six NEs there, only two are strict, namely

(25) a. $zzaz/pA$

b. $ezzz/pO$

For the other four NEs, the expected utility is not the unique maximum in its row. Therefore these strategies are vulnerable to invasion from either of the two SNEs. So while the pure accusative system $zzaa$ and the pure ergative system $eezz$ give rise to Nash Equilibria, these equilibria are not evolutionarily stable. For the scenarios with higher values of k, all NEs are strict and thus evolutionarily stable.

The situation for $nn > pp$ is essentially similar, except the subject marking and object marking are to be exchanged. For small values of k, split ergativity $zeaz/pA$ is the only SNE. If k grows larger, the SNEs $zezz/pA$ (DSM) and $zzza/pO$ (inverse DOM) coexist. Still higher values of k lead to the coexistence of $zzaz/pA$ (DOM) with $ezzz/pO$ (inverse DSM) and of $zzzz/pA$ (zero marking) with $ezzz/pO$, and finally, for very high costs, $zzzz/pA$ is the only SNE again.

As mentioned above, a prominence split which is high on some prominence hierarchy—like the split between local persons and third person—is more likely to have $nn > pp$, while low split points make $pp > nn$ more probable. So the configuration of evolutionarily stable strategy pairs depends on two dimensions, the position of the split point and the costs of case-marking. This is schematically expressed in Figure 3.

[Figure 3 goes about here.]

4.5 Stochastic evolution The major conceptual advantage of EGT over rationalistic GT is the fact that the latter has to assume that the players are aware of the respective utilities of the strategies at their disposal, and that they are able to draw all logically valid conclusions from this knowledge—a totally unrealistic assumption in the context of linguistic application—while in EGT equilibria emerge independently from the knowledge of the agents. Empirically, the step from the rationalistic to the evolutionary model can be seen as a progress or not, depending on the perspective taken. While the pure ergative system (which is very rare if it exists at all—Basque might be a case in point, but there is some debate on this issue; see for instance Joseph et al. 1989[29]) and the pure accusative system constitute Nash Equilibria, they are not evolutionarily stable. Languages of such a type are very rare, and it might thus be a good thing to characterize them as unstable. On the other hand, at least Hungarian is an indisputable example of a $zzaa$ pattern, while EGT predicts that in the long run only stable systems survive. So the existence of the Hungarian pattern appears to be a puzzle, and working in the context of classical EGT, we would have to resort to the assumption that for some reason, Hungarian is still in a pre-stable state.

On the other hand, the EGT model still includes the typologically uncommon inverse DOM ($zzza$) and inverse DSM ($ezzz$) language type. So if the predictions of the EGT model are taken to express universal tendencies, they are too inclusive (failing to exclude inverse DOM and inverse DSM), while if they are supposed to express clear-cut universals, they are too strong because they exclude existing languages like Hungarian. In this section, it is argued that a natural refinement of EGT, namely stochastic EGT, clarifies these issues. According to this approach, evolutionary stability can only be a gradient notion and the empirical predictions of an evolutionary approach are necessarily probabilistic rather than categorical. Furthermore, it is demonstrated that inverse DOM ($zzza/pO$) and inverse DSM ($ezzz/pO$) are not stable under the stochastic refinement.[30]

Consider the scenario where $pp > nn$ and $\frac{nn}{2(np+nn)} < k < 0.5$. Then we have two SNE, namely $zzaz/pA$ and $ezzz/pO$ (see Table 8). Since these are the only evolutionarily stable states, we may restrict attention to speaker populations that consist only of $zzaz$-players and $ezzz$-players, and hearers which play either pA or pO. The utility matrix for this simplified game is as in Table 12.

[Table 12 goes about here.]

Filling in the concrete values from Table 8 for illustration gives the speaker utilities as in Table 13.

[Table 13 goes about here.]

There is a strong asymmetry between the two SNE in two respects. First, the payoffs of both speaker and hearer are much higher in the SNE $zzaz/pA$ than in $ezzz/pO$.[31] In a rationalistic setting, this may be an incentive for rational players to settle for this equilibrium, but of course this is not an argument in an evolutionary setting. However, consider the invasion barriers of the two equilibria. Suppose the system is in the state $zzaz/pA$. Then at least 54.8% of all hearers have to mutate to the strategy pO before it is beneficial for the speaker to switch to $ezzz$ (or, in evolutionary terms, before a speaker mutation of the $ezzz$ would survive and drive $zzaz$ to extinction).[32] Even more dramatic, no less than 97.9% of all speakers would have to mutate to $ezzz$ before a hearer mutation to pO would survive. If the system is in the other equilibrium, $ezzz/pO$, things are reversed. 45.8% hearer mutations are needed to leave the basin of attraction of $ezzz/pO$, but already 2.1% of speaker mutations to $zzaz$, combined with an arbitrarily small amount of hearer mutations to pA, pushes the system into the $zzaz/pA$ state. So $zzaz/pA$ is much more robust against mutations than $ezzz/pO$.[33]

Let us now have a closer look at the modeling of mutations in EGT. The dynamics underlying the vector field in Figure 2 above contain a low intensity but constant stream of mutations. This is actually an artifact of the assumption that populations are infinite and time is continuous in standard EGT. Real populations are finite though, and both games and mutations are discrete events in time. So a more fine-grained modeling should assume finite populations and discrete time. Now suppose that for each individual in a population, the probability to mutate towards the strategy s within one time unit is p, where p may be very small but > 0. If the population consists of n individuals, the chance that all individuals end up playing s at a given point in time is at least p^n, which may be extremely small but is still positive. By the same kind of reasoning, it follows that there is a positive probability for a finite population to jump from each state to each other state due to mutation (provided each strategy can be the target of mutation of each other strategy). More generally, in a finite population the stream of mutations is not constant but noisy and nondeterministic. Hence there are strictly speaking no evolutionarily stable strategies because every invasion barrier will eventually be overcome, no matter how low the average mutation probability and how high the barrier is.[34]

If an asymmetric game has exactly two SNEs, A and B, in a finite population with mutations there is a positive probability p_{AB} that the system moves from A to B due to noisy mutation, and a probability p_{BA} for the reverse direction. If $p_{AB} > p_{BA}$, the former change will on average occur more often than the latter, and in the long run the population will spend more time in state B than in state A. Put differently, if such a system is observed at some arbitrary time, the probability that it is in state B is higher than that it is in A. The exact value of this probability converges towards $\frac{p_{AB}}{p_{AB}+p_{BA}}$ as time grows to infinity.

If the level of mutation gets smaller, both p_{AB} and p_{BA} get smaller, but at a different pace. p_{BA} approaches 0 much faster than p_{AB}, and thus $\frac{p_{AB}}{p_{AB}+p_{BA}}$ (and thus the probability of the system being in state B) converges to 1 as the mutation rate converges to 0. So while there is always a positive probability that the system is in state A, this probability can become arbitrarily small. A state is called stochastically stable if its probability converges to a value > 0 as the mutation rate approaches 0. In the described scenario, B would be the only stochastically stable state, while both A and B are evolutionarily stable. The notion of stochastic stability is a strengthening of the concept of evolutionary stability; every stochastically stable state is also evolutionarily stable,[35] but not the other way round.

Returning to our example, it seems intuitively obvious that a change from $ezzz/pO$ to $zzaz/pA$ is much more likely than the change in the reverse direction, and that hence DOM, but not inverse DSM is stochastically stable. There is a considerable body of literature, mainly in economics, on the mathematics of stochastic evolution.[36] Unfortunately though, there are no general and easy-to-use recipes for computing stochastic stability like Selten's identification of evolutionary stability with SNEs. It is possible to compute the stochastically stable states of a game analytically, but the required mathematics is quite involved and goes beyond the scope of the present article. However, it is fairly straightforward to demonstrate (if not to prove) that DOM but not inverse DSM is stochastically stable by using computer simulations, and the same applies ceteris paribus to inverse DOM.

I ran several simulations using a discrete time version of the replicator dynamics (which, as mentioned above, can formally be seen as a special case of imitation dynamics), augmented with a stochastic mutation component.[37] Depending on various parameters like population size and average mutation rate there are three qualitatively different kinds of behavior. If the mutation rate is very high, all strategies are more or less equally likely in the long run, and there is a high variance in the probabilities of the different strategies. If the mutation rate is reduced and the population size increased, the risk-dominant SNE $zzaz/pA$ emerges as relatively stable, while the second SNE, $ezzz/pO$ is as unstable as the strategy pairs that are not SNE. Finally, if noise is further reduced (i.e. population size grows and mutation rate shrinks), the second SNE becomes stable as well. We do find exactly the asymmetry between the two SNE though that was described before in theoretical terms. As long as the noise is not too small, a population in the $ezzz/pO$-state will soon or later switch into the $zzaz/pA$-state. The reverse change is theoretically possible as well, but its probability is so low that I never actually observed it, even in very long simulations. Figure 4 below shows the plots of a typical run of an experiment where evolution started in a pure $ezzz/pO$-population. The first plot shows the development of the two relevant speaker strategies, and the second one the two hearer strategies. The initial strategy combination remains dominant for some time, but the equilibrium is precarious and the variance is high. After some time the system shifts into the state $zzaz/pA$ and remains there for the remainder of the experiment. Here the equilibrium appears to be much more robust, especially for the hearer strategies the variance is significantly lower than in the initial equilibrium.

[Figure 4 goes about here.]

For all practical purposes one can say that with low mutation noise, both evolutionarily stable states are stable under stochastic evolution, but only $zzaz/pA$ is attainable. A system may remain in the state $ezzz/pO$ for some time, but it is virtually impossible to reach that state from any other state outside its basin of attraction. I found the same asymmetry for all instances of the Game of Case where two SNEs coexists. Inverse DOM and inverse DSM—i.e. the SNEs that involve the hearer strategy pO—are never stochastically stable. So under stochastic evolution, there are only four case-marking patterns that are stochastically stable: split ergative ($zeaz$), DOM ($zzaz$), DSM ($zezz$), and zero marking ($zzzz$). However, it follows from the very concept of stochastic mutation that each state has a positive probability. Hence the prediction of the present account is not that only these four case-marking patterns exist, but only that the vast majority of all languages will have one of these four types.

The configuration of stochastically stable states, depending on the position of the prominence split and on the cost coefficient, is given in Figure 5.

[Figure 5 goes about here.]

4.6 Desiderata So while the majority of languages are in a stochastically stable state, there are some exceptions. Most notably, nondifferential accusative systems $zzaa$ like Hungarian exist even though they are predicted to be stochastically unstable.

It is of course possible that such language are genuine exceptions. The typological predictions derivable from the evolutionary model are statistical in nature, not categorical. It is also possible though that this misprediction points to a relevant factor that has been ignored in the present model. It is remarkable that the pattern in question, $zzaa$, is among the simplest ones imaginable. The present model only takes selection pressure for phonological reduction and disambiguation into account. There is arguably also a selection pressure for learnability, and a system like $zzaa$ scores higher in this respect than a comparatively complex one like $zeaz$. If these competing motivations are taken into account, $zzaa$ may actually turn out to be stochastically stable. Future work will have to show whether speculation can be substantiated.

Also, the theory sketched here fails to predict the massive asymmetry between accusative and ergative languages. Accusative languages are much more common than ergative languages, and even morphologically ergative languages mostly have syntactic features of accusative languages. Why is this so?

A popular explanation follows Du Bois (1987) in the assumption that both the nominative/accusative split and the ergative/absolutive split have a basis in information structural dichotomies. According to this view, nominative NPs are predominantly topical and accusatives nontopical, while absolutive NPs are the preferred syntactic vehicles for introducing novel discourse entities, and ergative NPs mostly refer to familiar material. If one grants that the topic/comment distinction is more salient than the novel/familiar one, the dominance of accusative systems can be related to this fact.

This line of reasoning offers a very convincing explanation of the fact that almost all agreement systems are accusative-based. It is well known that at least diachronically, agreement markers frequently (but not always) arise from bound pronouns that pick up the referent of the sentence topic (see for instance Givon 1976:151). However, I am not sure that a similar explanation carries over to the asymmetry between morphological accusativity vs. ergativity. Almost all accusative languages have DOM. This means that accusative marking is confined to the upper segment of the definiteness or animacy scale, and in either case the high segments of these scales covers those NPs which are most likely topical. So a good case could be made that accusative case, rather than nominative case, is correlated to topicality. In other words, there is no natural information structural dichotomy which would separate prominent transitive objects on the one hand from subjects and nonprominent objects on the other hand.

The game-theoretic analysis of case-marking suggests an alternative explanation. DOM languages confine accusative marking to the upper segment of the definiteness/animacy scale. In terms of NP tokens, a substantial proportion of prominent objects are pronouns, i.e. closed class elements, even if the split point is low. DSM, however, means that ergative marking is confined to lower segments of the scale. These cover mostly or even exclusively open class elements. Closed class items are more frequent than open class items, and it is a well-known fact that frequent forms tend to preserve more distinctions than rare ones.[38] In accusative languages this tendency and the tendency to restrict case-marking to an upper segment of the prominence scale coincide. In ergative languages, the tendency to restrict case-marking to the less prominent elements is opposed to the tendency induced by this frequency effect. This fact might be responsible for the comparable rarity of ergative languages.

5 Discussion In this section some questions are taken up that have repeatedly come up in discussions of the material from this article, as well as in the comments of the reviewers and editors of Language.

- Your model makes strong idealizations. For instance, it ignores word order, agreement, information structure etc., which are all relevant for argument linking and case-marking. What can we possibly learn from such an oversimplified model?

  Simplification and idealization can be good or bad features of a model, depending what the model is supposed to do. If one tries to implement a flight simulator to train pilots, say, it is advisable to model as many aspects of reality as possible, to make the model as realistic as possible. The purpose of a scientific model, however, is not to approximate reality as closely as possible, but to explore the consequences of theoretical assumptions and to generate empirically testable hypotheses. To take an example from physics, classical mechanics models rigid bodies as mass points. This is a gross oversimplification, because it abstracts away from factors like shape and density. Nonetheless the classical model is highly successful in modeling inertia and gravitation. The simplified model thus provided the insight that shape and density are irrelevant for inertia and gravitation. The fact that the classical model fails to predict for instance ballistic curves correctly is also important; from this we learn that one of the factors that the simple model ignores is relevant here—in this example shape, in connection with air resistance.

  In the present paper I have attempted to establish a correlation between two empirical domains, the relative frequencies of transitive clause types in performance on the one hand, and the typological distribution of case-marking patterns on the other hand. The crucial theoretical assumptions connecting those domains are

  - that grammars are self-replicating systems, and
  - that the success of replication is negatively correlated with both markedness and with ambiguity (meaning: grammars that use marked constructions and lead to ambiguity are less successful in replication than grammars that avoid marked constructions and encode meanings unambiguously).

Many predictions that follow from these assumptions are in fact supported by the empirical data. This is evidence, if not proof, that the assumptions are on the right track, and that the statistical patterns in performance are in fact the main factor that is responsible for the correlation between differential case-marking and the nominal hierarchies. On the other hand, these assumptions fail to predict certain salient facts, like the typological asymmetry between accusative and ergative languages (see the discussion above). This indicates that we have to look out for other relevant factors here.

- What exactly do you mean by 'self-replication'? Can such biological notions really be extrapolated to linguistics in such a straightforward way?

  As pointed out in section 4.4, EGT as a mathematical framework is compatible with several conceptualizations of language evolution. Let me sketch what I consider to be a plausible basic mechanism. It is crucially inspired by Croft 2000, but there are also significant differences.

  A unit of linguistic replication is 'any piece of structure that can be independently learned and therefore transmitted from one speaker to another' (Nettle 1999:5). Nettle calls these units 'items', while Croft uses the term 'linguemes' (that he attributes to Martin Haspelmath). A lingueme in this sense can be a phoneme, a morpheme, a word, an idiom, a syntactic construction, what have you. Linguemes can be replicated in at least two ways:

  – In (first or second) language acquisition, the learners store linguemes in their memory that they pick up from their environment, and they retrieve those items later in language usage.
  – Speakers tend to repeat linguemes that they just encountered. This process is called 'priming' in psycholinguistics (Bock 1986, Bock et al. 1992, Zwitserlood 1996, Branigan et al. 2000, Pickering and Branigan 1999). Priming is bidirectional: both comprehension and production can trigger priming. (In the

first case one can speak about imitation, in the second about repetition or self-imitation.)

The first kind of replication is the primary one. Language users use linguemes primarily to communicate, and they use the linguemes that they learned and that serve their communicative purposes. Furthermore, priming is possible only if the individual that gets primed has acquired the lingueme in question before. So priming can be seen as a kind of filter or amplifier that modifies the effects of the primary mode of replication. Metaphorically speaking, the usage of a lingueme in an utterance has thus many ‘parents’—every previous encounter of the speaker with this lingueme that is causally related to its present use.

Linguemes are the replicators in this view on language evolution, corresponding to genes in biology. An utterance is a structured entity, composed of many linguemes. To the extent that there is an analogy between utterances in linguistics and genotypes in biology (see Croft 2000:12). According to Croft, the linguistic counterpart of the phenotype is the speaker. I would sharpen this further: the phenotypes, i.e. the interactors on which selection operates, are language-users-in-an-utterance. Language users may acquire or lose linguemes, and they are constantly re-shaped by priming. Therefore they constantly change the properties that are relevant for language evolution.

An essential precondition for any kind of evolution is variation in replication. In biology this is ensured by mutations. There is no direct counterpart to mutations in linguistic evolution, but there are sources of variation nevertheless. First, priming is not necessarily reflexive.[39] Processing a given lingueme does not only make the re-use of the very same lingueme more likely, but it also increases the likelihood of the usage of related linguemes. This relatedness can be semantic or formal. Significantly, priming can be asymmetric. For instance, Shields and Balota (1991) show experimentally that in repetition priming, the prime target is phonetically reduced in comparison to the prime source. Together with reanalysis, repeated phonetic reduction can lead to phonological reduction, i.e. the creation of new linguemes.

There may also be a difference between prime source and prime target in their semantics and grammatical status. Bock and Loebell (1990) for instance show that prepositional locatives (‘The wealthy woman drove the Mercedes to the church.’) can prime prepositional datives (‘The boy is giving a guitar to the singer.’). Likewise, intransitive locatives (‘The 747 was landing by the control tower’) can prime passive constructions (‘The 747 was alerted by the control tower’). Here too, nonreflexive priming and reanalysis can jointly lead to the creation of new linguemes.

Another source of new linguemes, which is not a consequence of imperfect replication, is ordinary linguistic creativity. Every complex linguistic structure that can be grammatically constructed out of several linguemes can be hardened into a new lingueme, as soon as somebody stores it cognitively as an integral unit. There is no real counterpart to such a source of variation in biology.

These remarks on variation barely scratch the surface of this very complicated issue, and this is not the right place to go deeper into it. With regard to the evolution of case-marking systems, the only kind of variation that matters is the acquisition and loss of case-marking. The latter can be the result of phonological reduction, i.e. asymmetric priming. Recruitment of other linguistic devices (like local prepositions) for case-marking purposes can be the result of reanalysis and subsequent phonological reduction.

Finally, variation leads to differential replication. Some replicators (genes, linguemes) are more likely to be replicated than others. This leads to selection. Croft claims that selection is mainly social in nature—those linguemes get preferably replicated that are socially useful for the language users. This is undoubtedly true. However, some linguemes are more apt to trigger priming than others, and successful primers will, all else being equal, be more successful replicators in general. Unlike Croft, I thus assume that there is a secondary selection pressure towards primability.

How does this conceptualization of language evolution relate to EGT? A strategy corresponds to an individual language user's grammar. This is a collection of linguemes, paired with a (possibly probabilistic) disposition to use certain linguemes in certain situations. The average replication probabilities of the linguemes in a grammar jointly determine the replication probability of the grammar as a whole. High replication probability of a grammar/strategy translates into a high utility (and vice versa). Whether or not the usage of a lingueme in an utterance causes its replication depends on the disposition (grammar/strategy) of the hearer of the utterance.[40] The particular utility function I used above implements essentially two claims:

- nominative has a higher utility than both accusative and ergative, and
- correct argument linking (in the sense that the hearer decodes the message of the speaker correctly) yields a higher utility than misunderstanding.

If utility is a measure of replicative success, these are claims about the differential replicative success of different linguemes. The first assumption basically translates into the hypothesis that nominative forms have on average a higher probability to be primed than accusative or ergative forms. This seems plausible, given that occurrences of nominative in total outnumber occurrences of any other case.[41] Frequent items are primed more often and thus have a higher average activation level than rare items. This justifies the claim that the utility of nominative forms is on average higher than the utility of other case forms. The above claim is actually stronger because it entails that this holds for every individual NP (type, not token). To test this hypothesis is well beyond the scope of this paper. However, in the absence of evidence to the contrary, the null hypothesis is that each individual NP behaves like average NPs.

The fact that nominative forms are usually shorter/less marked than other case forms, and the assumption that nominative yields a higher utility than other cases are thus both consequences of the same frequency bias. The utility of an item is correlated with markedness, but the former cannot be reduced to the latter. So even in languages with marked nominative, it is justified to assume that nominative has a higher utility than accusative/ergative.

Let us now turn to the second component of the utility function. Suppose a speaker wants to express a certain meaning m, and she uses the form f to communicate m. If communication succeeds, the hearer maps f back to m.

Elementary associations between forms and meanings are linguemes. The speaker's usage of the lingueme $m - f$ (or the elementary form-meaning mappings that this association is composed of) primes this very lingueme for the speaker. Due to the bidirectional nature of priming, it also primes that lingueme for the hearer; the disposition of the hearer to express the meaning m as f is strengthened, as is his disposition to interpret f as m on next encounter. So the speaker's usage of the lingueme $m - f$ replicates this lingueme in two ways—via self-priming and via bidirectional priming of the hearer.

Now suppose communication fails; the hearer interprets f as some meaning m' that is different from m. The speaker's self-priming is unaffected by this, but the utterance of the speaker fails to prime the lingueme $m - f$ for the hearer. So the replication probability of this lingueme from the speaker perspective is much lower than in the case of successful communication. The hearer's usage of the association $m' - f$ will cause self-priming; so the replication of the hearer's strategy seems to be unimpeded by the misunderstanding. However, due to the redundancy of linguistic communication, the meaning m' is likely to be incoherent or contradictory. It seems plausible that this weakens the hearer's self-priming.

If these considerations are correct, successful communication leads to a higher replication probability of the linguemes involved than failed communication. Expressed in terms of EGT, this means that successful communication has a higher utility than misunderstanding.

I hasten to point out that all these claims are speculative, but they are empirically testable. Psycholinguistic research may show them to be false or in need of refinement. The more general point to be made here is that EGT is a framework that enables us to connect the psycholinguistic micro-structure of linguistic replication with its large-scale typological consequences in a mathematically precise way.

- Biological evolution is fundamentally different from linguistic evolution because biological mutations are random and undirected, while the sources of linguistic variation are directed and adaptive. Therefore the biological notion of natural selection cannot be applied to linguistics. In general this is a valid and important point. For the particular analysis that was developed in this paper, the directedness of linguistic variation is not such a severe a problem though.

  The notion of evolutionary stability is very general. It does not require that all mutations are equally likely. To apply this notion, it is sufficient that there is a 'mutation' path from each strategy to each other strategy. In the application at hand, this means that there must be diachronic paths that neutralize a nominative/accusative distinction, and diachronic paths that establish such a distinction; and likewise for nominative/ergative. There is ample evidence from historical linguistics for these processes. For instance, Latin had differential object marking ('$zzza$'), with a nominative/accusative distinction for masculine and feminine, but not neuter. This distinction was lost in the first centuries C.E. via phonological erosion (Blake 2001:177). Much later, in Spanish a novel accusative marker (the preposition a) emerged, a generalization of an indirect object marker, which in turn derives from the Latin preposition ad 'to' (see Blake 2001:173). This new accusative marker is confined to animate objects. So the pattern of differential object marking is actually reestablished, if based on a different split criterion (animacy instead of grammatical gender).

  We do not have comparable historical records for the loss or emergence of ergative systems, but that these pathways exist has been demonstrated by comparative methods. For instance, an accusative system can be transformed into an ergative system via re-analysis of the passive construction, and change in the opposite direction may be due to re-analysis of anti-passive (see the discussion in Dixon 1994, ch. 7).

The notion of stochastic stability (as opposed to mere evolutionary stability) is based on the assumption that all mutations/pathways are equally likely. Here again, the null hypothesis in the absence of precise information is to assume a uniform probability distribution, but this may have to be modified if it can be shown to lead to wrong conclusions.

- One central notion of the paper is 'stability' ('evolutionary' or 'stochastic'). How long must a case system remain constant to qualify as stable? Are there any truly (diachronically) stable linguistic phenomena at all? Do you have diachronic data to support your claims about stability? It is important to appreciate that the two notions of stability that are used in the analysis—evolutionary stability and stochastic stability—are theoretical notions, not empirical ones. A state is evolutionarily stable if a system does not leave it provided the mutation rate is arbitrarily small. A state is stochastically stable if its probability converges to a positive value when the mutation rate converges to zero. The true, empirical mutation rate is not arbitrarily small, and it does not converge to zero. The definitions of these notions of stability thus invoke contra-factual statements that cannot directly be tested empirically. This does not devalue them as parts of an analysis—that a grain of salt is soluble in water is ultimately a fact about its chemical composition, and as such it can be true or false, even if the grain never gets in contact with water. Likewise, the two notions of stability ultimately translate into statements about the role of states in diachronic change. Briefly put, a state is stochastically stable if it is more likely to be the target than the source of language change. The results about stochastic stability are thus empirical claims about diachrony. I haven't been able to test them so far, and, frankly, I am actually skeptical that this is practically possible at all. However, they lead to predictions about the synchronic typological distribution of states. A language is more likely to be in a stochastically stable state than in an unstable state. According to the laws of probability, if the languages of the world fulfill the minimal standards for a statistical sample, we expect that the majority of them are in a stochastically stable state. As I have tried to argue here, this prediction is borne out by the findings of typological research.

References

Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory* 17.673–711.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21.435–483.

Blake, Barry J. 2001. *Case*. Cambridge: Cambridge University Press.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18.355–387.

Bock, J. Kathryn and Helga Loebell. 1990. Framing sentences. *Cognition* 35.1–39.

Bock, J. Kathryn, Helga Loebell, and Randal Morey. 1992. From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological Review* 99.150–171.

Bossong, Georg. 1985. *Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Günther Narr Verlag.

Branigan, Holly P., Martin J. Pickering, Andrew J. Stewart, and Janet F. McLean. 2000. Syntactic priming in spoken production: linguistic and temporal interference. *Memory & Cognition* 28.1297–1302.

Breen, Gavan. 1976. Ergative, locative and instrumental inflections in Wangkumara. *Grammatical categories in Australian languages*, ed. by R. M. W. Dixon, 336–339, Canberra: Australian Institute of Aboriginal Studies.

Bybee, Joan and Sandra Thompson. 2000. Three frequency effects in syntax. *Berkeley Linguistic Society* 23.65-85.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Comrie, Bernard. 1979. Definite and animate direct objects: A natural class. *Linguistica Silesiana* 3.13–21.

Comrie, Bernard. 1981. *Language universals and linguistic typology*. Oxford: Basil Blackwell.

Croft, William. 2000. *Explaining language change*. New York: Longman.

Dahl, Östen. 2000. Egophoricity in discourse and syntax. *Functions of language* 7.37–77.

Dixon, R. M. W. 1972. *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.

Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Du Bois, John. 1987. The discourse base of ergativity. *Language* 63.805–855.

Fisher, Ronald Aylmer. 1930. *The genetical theory of natural selection*. Oxford: Clarendon Press.

Givon, Talmy. 1976. Topic, Pronoun, and Grammatical Agreement. *Subject and Topic*, ed. by Charles N. Li, 149–188, New York: Academic Press.

Greenberg, Joseph. 1963. Some universals of grammar with special reference to the order of meaningful elements. *Universals of language*, ed. by Joseph Greenberg, 73–113, Cambridge: MIT Press.

Haspelmath, Martin. 1999. Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18.180–205.

Hofbauer, Josef and Karl Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.

Jäger, Gerhard and Anette Rosenbach. 2005. Priming as a driving force in grammaticalization: on the track of unidirectionality, paper presented at *New Reflections on Grammaticalization 3*, Santiago de Compostela.

Jäger, Gerhard and Robert van Rooij. 2005. Language stucture: Psychological and social constraints. manuscript, University of Bielefeld and University of Amsterdam.

Joseph, Brian D., Kutz Arrieta, and Jane Smirniotopoulos. 1989. How Ergative is Basque? *Proceedings of the Third Eastern States Conference on Linguistics [ESCOL '86]*, 25–36, Columbus: OSU Department of Linguistics.

Kandori, Michihiro, George Mailath, and Rafael Rob. 1993. Learning, mutation, and long-run equilibria in games. *Econometrica* 61.29–56.

Kirby, Simon. 1999. *Function, selection, and innateness*. Oxford: Oxford University Press.

Kroch, Anthony. 2000. Syntactic change. *The handbook of contemporary syntactic theory*, ed. by Mark Baltin and Chris Collins, 699–729, Oxford: Blackwell Publishers.

Kumakhov, Mukhadin, Karina Vamling, and Zara Kumakhova. 1996. Ergative case in the Circassian languages. Working Papers 45, Lund: Lund University, Dept. of Linguistics.

Lewis, David. 1969. *Convention*. Cambridge: Harvard University Press.

Löfström, Jonas. 1988. Repliker utan gränser. till studiet av syntaktisk struktur i samtal. Gothenburg: Institutionen för nordiska språk, Göteborgs universitet.

Maynard Smith, John. 1982. *Evolution and the theory of games*. Cambridge: Cambridge University Press.

Næss, Åshild. 2004. What markedness marks: The markedness problem with direct objects. *Lingua* 114.1186–1212.

Nettle, Daniel. 1999. *Linguistic diversity*. Oxford: Oxford University Press.

Nilsenova, Marie. 2002. The Pragmatics of Differential Object Marking. *Proceedings of the ESSLLI '02 Student Session*, ed. by Malvina Nissim, Trento: Folli.

Nowak, Martin A., Natalia L. Komarova, and Partha Niyogi. 2002. Computational and evolutionary aspects of language. *Nature* 417.611–617.

Osborne, Martin J. 2003. *An Introduction to game theory*. Oxford: Oxford University Press.

Parikh, Prashant. 2001. *The use of language*. Stanford: CSLI Publications.

Pickering, Martin J. and Holly P. Branigan. 1999. Syntactic priming in language production. *Trends in Cognitive Sciences* 3.136–141.

Plank, Frans. 2001. Das grammatische Raritätenkabinett. manuscript, University of Konstanz, available from `www.ling.uni-konstanz.de/~pages/~proj/~Sprachbau/~rara.html`.

Reed, I., O. Miyaoka, S. Jacobsen, P. Afcan, and M. Krauss. 1977. *Yup'ik Eskimo grammar*. University of Alaska.

Selten, Reinhard. 1980. A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology* 84.93–101.

Shields, Lynn W. and David A. Balota. 1991. Repetition and associative context effects in speech production. *Language and Speech* 34.47–55.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. *Grammatical categories in Australian languages*, ed. by R. M. W. Dixon, 112–171, Canberra: Australian Institute of Aboriginal Studies.

van Rooij, Robert. 2004. Signalling games select Horn strategies. *Linguistics and Philosophy* 27.493–527.

Vega-Redondo, Fernando. 1996. *Evolution, games, and economic Behaviour*. Oxford: Oxford University Press.

von Neumann, John and Oskar Morgenstern. 1944. *Theory of games and economic behavior*. Princeton: John Wiley & Sons.

Young, H. Peyton. 1993. The evolution of conventions. *Econometrica* 61.57–84.

Young, H. Peyton. 1998. *Individual strategy and social structure. An evolutionary theory of institutions*. Princeton: Princeton University Press.

Zeevat, Henk and Gerhard Jäger. 2002. A reinterpretation of syntactic alignment. *Proceedings of the Fourth International Tbilisi Symposium on Language, Logic and Computation*, ed. by Dick de Jongh, Marie Nielsenova, and Henk Zeevat, Amsterdam: University of Amsterdam.

Zipf, G. K. 1949. *Human behavior and the principle of least effort*. Cambridge: Addison Wesley.

Zwitserlood, Pienie. 1996. Form priming. *Language and Cognitive Processes* 11.589–596.

Notes

[1]See for instance Comrie (1979).

[2]By this I mean the partial order over the Cartesian product of the domain of the two scales, where $\langle a_1, b_1\rangle \geq \langle a_2, b_2\rangle$ iff $a_1 \geq a_2$ and $b_1 \geq b_2$.

[3]Here and henceforth, I use the term 'subject' to refer both to the single argument of an intransitive verb and to the controller/agent argument of a transitive verb. 'Object' refers to

the nonsubject argument of a simple transitive verb. While this terminology expresses a bias towards accusative systems and against ergative systems, no real harm is done in the context of this paper because it does not deal with intransitive clauses.

[4]As Comrie (1981:118) puts it: 'The last type, with A/P–S alignment, seems to be equally rare: the only reliable attestations known to us are for certain classes of noun phrases in certain Iranian languages, where it represents an intermediate diachronic stage in the breakdown of an earlier ergative-absolutive system case-marking system in the direction of a nominative-accusative system.'

[5]Following common practice, the term 'nominative' is used from now on to designate the case that is used for 'S', thus subsuming the traditional notion of 'absolutive' under 'nominative'.

[6]The question of the proper interpretation of this principle is addressed later on, but let us take it for granted for now.

[7]See `http://www.grsampson.net/RChristine.html`

[8]This corpus is a subset of the corpus 'Samtal i Göteborg' (Conversations in Göteborg); see Löfström 1988. The subset that I used (about 60,000 words) has been hand-annotated by Östen Dahl for syntactic roles and animacy (see Dahl 2000). I thank Östen Dahl for the kind permission to use his data.

[9]In the game-theoretic literature, this kind of game is called signaling game. This class of games was originally introduced by Lewis (1969) in the context of his study of conventions. Recent applications of this model to linguistic problems are Parikh (2001) and van Rooij (2004).

[10]In a more realistic model, the distance between two meanings should be measurable in a multi-dimensional continuous vector space; see Jäger and van Rooij 2005 for a generalization of the present model along those lines. In the context of this paper, we consider only the choice between correct and incorrect argument linking, which is in fact a binary choice.

[11]The notions of ‘payoff’ and ‘utility’ are interchangeable.

[12]I consider only unmarked nominative systems in what follows, and in those systems neither speaker nor hearer have a choice which could be modeled by means of strategies. Non-core functions like indirect object, possessor etc. are not treated here.

[13]If one of the two NPs is elided, I assume that the elided phrase precedes the overt one, for the sake of the argument. Nothing depends on this.

[14]It would perhaps be more precise to speak of eight meaning types, but I trust no confusion arises from this.

[15]It is well known that in languages without articles, case-marking may be an indicator of (in)definiteness of an NP. To keep things simple, I deliberately exclude such a semantic overloading of case morphology.

[16]Again, there are languages which are more complex than that—I return to that point later on.

[17]The classical reference is von Neumann and Morgenstern 1944; a fairly accessible introduction can be found in Osborne 2003.

[18]One reviewer points out that there are languages, which employ a split case system which depends on tense, rather than on the semantics of the NPs involved. In Hindi, for instance, ‘present tense clause follow a nominative/accusative case-marking pattern but past tense clauses follow an ergative scheme’. In the present model, the different tenses of Hindi correspond to different strategies.

[19]In GT terminology, the Game of Case thus becomes a partnership game because all players receive the same payoff in each game.

[20]I am actually not aware of a single undisputed instance of this pattern. Næss (2004) mentions Mayan as a possible example. However, ergative marking is confined to pronouns there,

and these pronouns could as well be analyzed as agreement markers.

[21]If k equals one of these thresholds, all NEs of the neighboring segments are NEs.

[22]This includes languages where the split is induced by grammatical gender, as it is the case in many Indo-European languages. As Blake points out a few sentences later: ‘In Latin and the other Indo-European case languages, there is no nominative/accusative distinction with neuter nouns. This is related to animacy in that virtually all neuter nouns are inanimate, though inanimate nouns are also plentiful in the masculine and feminine genders.’

[23]Replication is to be thought of as asexual, i.e. each individual has exactly one parent.

[24]In the simplest model of EGT, based on the so-called replicator dynamics, populations are—

simplifyingly—thought of as infinite and continuous, so there are no minimal units.

[25]See for instance Hofbauer and Sigmund 1998:86.

[26]Of course the same individual can be both speaker and hearer at different occasions. As long as the speaker strategy and the hearer strategy of an individual do not influence each other, this is formally still equivalent to a two-population model. Needless to say, this independence assumption is a strong idealization.

[27]In symmetric games the situation is slightly more complex because a mutant will occasionally encounter other mutants of the same type, and evolutionary stability also depends on how well a mutant strategy does against itself. In multi-population games, a mutant strategy will never play against itself simply because the participants of a game belong to different populations.

[28]Actually the game is larger because it also comprises six speaker strategies that are strictly dominated. However, strictly dominated strategies always die out in a population dynamic setting and can thus safely be ignored.

[29]Thanks to Brian Joseph for drawing my attention to this point and this reference.

[30]The first application of stochastic EGT, and of EGT in general, to linguistic problems is van Rooij 2004, where the emergence of iconicity is investigated. Van Rooij shows that both iconic and anti-iconic patterns are evolutionarily stable, but only the former are also stochastically stable.

[31]In the game-theoretic terminology, $zzaz/pA$ is the Pareto-efficient SNE in this game.

[32]Here is how the figure 54.8$\%$ is calculated: Suppose the proportion of hearers that play pA is p. Then the utility of the incumbent speaker strategy $zzaz$ is

$$p \times u_s(zzaz, pA) + (1 - p) \times u_s(zzaz, pO)$$

Likewise, the utility of $ezzz$ is

$$p \times u_s(ezzz, pA) + (1 - p) \times u_s(ezzz, pO)$$

Filling in the numerical values from Table 13, it turns out that $zzaz$ has a higher utility iff $p < .548$, and $ezzz$ has a higher utility if $p > .548$.

[33]In the rationalistic terminology, $zzaz/pA$ is risk-dominant. If you are a perfectly rational player but you cannot be absolutely sure that the other player is rational, it is advisable to play the risk-dominant Nash Equilibrium if there is one. In the games considered here, risk-dominance and Pareto-efficiency always coincide.

[34]This idea was first developed in Kandori et al. 1993 and Young 1993. Fairly accessible introductions to the theory of stochastic evolution are given in Vega-Redondo 1996 and Young 1998.

[35]Provided the population is sufficiently large, that is. Very small populations may display a weird dynamic behavior, but I skip over this side aspect here.

[36]See for instance Vega-Redondo 1996 or Young 1998 for overviews and further references.

[37]The system of difference equations used in the experiments is

$$\begin{aligned}\frac{\Delta x_i}{\Delta t} &= x_i((A\mathbf{y})_i - \langle \mathbf{x} \cdot A\mathbf{y} \rangle) + \sum_j \frac{Z_{ji} - Z_{ij}}{n} \\ \frac{\Delta y_i}{\Delta t} &= y_i((B\mathbf{x})_i - \langle \mathbf{y} \cdot B\mathbf{x} \rangle) + \sum_j \frac{Z_{ji} - Z_{ij}}{n}\end{aligned}$$

where $\mathbf{x}, \mathbf{y}$ are the vectors of the relative frequencies of the speaker strategies and hearer strategies, and A and B are the payoff matrices of speakers and hearers respectively. For each pair of strategies i and j belonging to the same player, Z_{ij} gives the number of individuals that mutate from i to j. Z_{ij} is a random variable which is distributed according to the binomial distribution $b(p_{ij}, \lfloor x_i n \rfloor)$ (or $b(p_{ij}, \lfloor y_i n \rfloor)$ respectively), where p_{ij} is the probability that an arbitrary individual of type i mutates to type j within one time unit, and n is the size of the population. I assumed that both populations have the same size.

[38]See for instance Bybee and Thompson (2000): ‘Pronouns and full NPs are related in the sense that pronouns diachronically derive from nouns, and synchronically, in that pronouns and NPs often occupy the same positions. However, a major difference between nouns and pronouns is that the latter are much more frequent than the former. This fact can be used as an explanation for why English pronouns maintain distinct forms for nominative vs. dative/accusative case, while nouns have lost these case distinctions.’

[39]The subsequent discussion about priming essentially follows Jäger and Rosenbach 2005.

[40]Modeling communication as an asymmetric signaling game might a bit seem artificial since it separates production grammar and comprehension grammar, even though both draw from the same lingueme pool in the mind of a single person. However, it is always possible to transform an asymmetric game into a symmetric game, and all results about evolutionary or stochastic stability remain valid under this transformation.

[41]At least in non-pro-drop languages, that is. In such a language, nominative is used in every clause, but all other cases only in some clauses. In pro-drop languages, the quantitative patterns may be different. The impact of pro-drop on frequency distributions, and thus on evolutionary stability, is an exciting issue for further research.

	Unmarked	Marked
Local persons	Subject	Object
3rd person	Object	Subject (of transitive)
Case	Nominative/Absolutive	Accusative/Ergative

Table 1: Case marking system of Dyirbal

	O/p	O/n
A/p	198	716
A/n	16	75

Table 2: Classification according to pronoun/full NP

pronoun/proper noun

local vs. 3rd person

	O/p	O/n
A/p	39	764
A/n	8	194

vs. complex NP

	O/p	O/n
A/p	255	698
A/n	11	41

definite vs. indefinite

	O/p	O/n
A/p	454	519
A/n	13	19

Table 3: Frequencies of clause types in CHRISTINE corpus

animate vs. inanimate

	O/p	O/n
A/p	300	2648
A/n	17	186

Table 4: Frequencies of clause types in corpus of spoken Swedish: animacy

definite vs. indefinite

	O/p	O/n
A/p	1806	1292
A/n	24	29

Table 5: Frequencies of clause types in corpus of spoken Swedish: animacy

Speaker strategies	Hearer strategies				
	AO	pA	pO	OA	$k \times$ costs
$eezz$	1	1	1	1	k
$zzaa$	1	1	1	1	k
$ezaz$	$1 - \frac{nn}{2N}$	$1 - \frac{nn}{2N}$	$1 - \frac{nn}{2N}$	$1 - \frac{nn}{2N}$	$k \times (1 - \frac{nn-pp}{N})$
$zeza$	$1 - \frac{pp}{2N}$	$1 - \frac{pp}{2N}$	$1 - \frac{pp}{2N}$	$1 - \frac{pp}{2N}$	$k \times (1 - \frac{pp-nn}{N})$
$zeaz$	$1 - \frac{pn}{2N}$	1	$1 - \frac{pn}{N}$	$1 - \frac{pn}{2N}$	$k \times (1 - \frac{pn-np}{N})$
$ezzz$	$1 - \frac{np+nn}{2N}$	$1 - \frac{2np+nn}{2N}$	$1 - \frac{nn}{2N}$	$1 - \frac{np+nn}{2N}$	$k \times \frac{pn+pp}{N}$
$zezz$	$1 - \frac{pn+pp}{2N}$	$1 - \frac{pp}{2N}$	$1 - \frac{2pn+pp}{2N}$	$1 - \frac{pn+pp}{2N}$	$k \times \frac{np+nn}{N}$
$zzaz$	$1 - \frac{pn+nn}{2N}$	$1 - \frac{nn}{2N}$	$1 - \frac{2pn+nn}{2N}$	$1 - \frac{pn+nn}{2N}$	$k \times \frac{pp+np}{N}$
$zzza$	$1 - \frac{np+pp}{2N}$	$1 - \frac{2np+pp}{2N}$	$1 - \frac{pp}{2N}$	$1 - \frac{np+pp}{2N}$	$k \times \frac{pn+nn}{N}$
$zzzz$	0.5	$1 - \frac{pp+2np+nn}{2N}$	$1 - \frac{pp+2pn+nn}{2N}$	0.5	0

Table 6: The Game of Case

Speaker strategies	Hearer strategies			
	AO	pA	pO	OA
$eezz$	0.90	0.90	0.90	0.90
$zzaa$	0.90	0.90	0.90	0.90
$ezaz$	0.85	0.85	0.85	0.85
$zeza$	0.81	0.81	0.81	0.81
$zeaz$	0.61	0.97	0.26	0.61
$ezzz$	0.86	0.86	0.87	0.86
$zezz$	0.54	0.89	0.54	0.54
$zzaz$	0.59	0.94	0.59	0.59
$zzza$	0.81	0.81	0.82	0.81
$zzzz$	0.50	0.85	0.15	0.50

Table 7: Frequencies: pronoun/full NP in CHRISTINE, $k = 0.1$

Speaker strategies	Hearer strategies			
	AO	pA	pO	OA
$eezz$	0.550	0.550	0.550	0.550
$zzaa$	0.550	0.550	0.550	0.550
$ezaz$	0.458	0.458	0.458	0.458
$zeza$	0.507	0.507	0.507	0.507
$zeaz$	0.507	0.863	0.151	0.507
$ezzz$	0.545	0.538	0.553	0.545
$zezz$	0.505	0.861	0.148	0.505
$zzaz$	0.510	0.867	0.154	0.510
$zzza$	0.539	0.531	0.547	0.539
$zzzz$	0.500	0.849	0.152	0.500

Table 8: Frequencies: pronoun/full NP in CHRISTINE, $k = 0.45$

Speaker strategies	Hearer strategies			
	AO	pA	pO	OA
$eezz$	0.470	0.470	0.470	0.470
$zzaa$	0.470	0.470	0.470	0.470
$ezaz$	0.368	0.368	0.368	0.368
$zeza$	0.436	0.436	0.436	0.436
$zeaz$	0.483	0.839	0.127	0.483
$ezzz$	0.473	0.465	0.480	0.473
$zezz$	0.497	$\boxed{0.854}$	0.141	0.497
$zzaz$	0.494	0.850	0.137	0.494
$zzza$	0.476	0.468	$\boxed{0.484}$	0.476
$zzzz$	0.500	0.848	0.152	0.500

Table 9: Frequencies: pronoun/full NP in CHRISTINE, $k = 0.53$

Speaker strategies	Hearer strategies			
	AO	pA	pO	OA
$eezz$	0.300	0.300	0.300	0.300
$zzaa$	0.300	0.300	0.300	0.300
$ezaz$	0.177	0.177	0.177	0.177
$zeza$	0.287	0.287	0.287	0.287
$zeaz$	0.431	0.788	0.075	0.431
$ezzz$	0.318	0.310	0.326	0.318
$zezz$	0.482	0.838	0.126	0.482
$zzaz$	0.457	0.814	0.101	0.457
$zzza$	0.343	0.335	$\boxed{0.350}$	0.343
$zzzz$	0.500	$\boxed{0.848}$	0.152	0.500

Table 10: Frequencies: pronoun/full NP in CHRISTINE, $k = 0.7$

Speaker strategies	Hearer strategies AO	pA	pO	OA
$eezz$	0.000	0.000	0.000	0.000
$zzaa$	0.000	0.000	0.000	0.000
$ezaz$	−0.160	−0.160	−0.160	−0.160
$zeza$	0.024	0.024	0.024	0.024
$zeaz$	0.340	0.697	−0.016	0.340
$ezzz$	0.045	0.037	0.053	0.045
$zezz$	0.455	0.811	0.099	0.455
$zzaz$	0.394	0.750	0.037	0.394
$zzza$	0.106	0.098	0.144	0.106
$zzzz$	0.500	$\boxed{0.848}$	0.152	0.500

Table 11: Frequencies: pronoun/full NP in CHRISTINE, $k = 1$

Speaker strategies	Hearer strategies pA	pO	$k \times$ costs
$ezzz$	$1 - \frac{2np+nn}{2}$	$1 - \frac{nn}{2}$	$k \times (pn + pp)$
$zzaz$	$1 - \frac{nn}{2}$	$1 - \frac{2pn+nn}{2}$	$k \times (pp + np)$

Table 12:

Speaker strategies	Hearer strategies	
	pA	pO
$ezzz$	0.538	0.553
$zzaz$	0.867	0.154

Table 13:

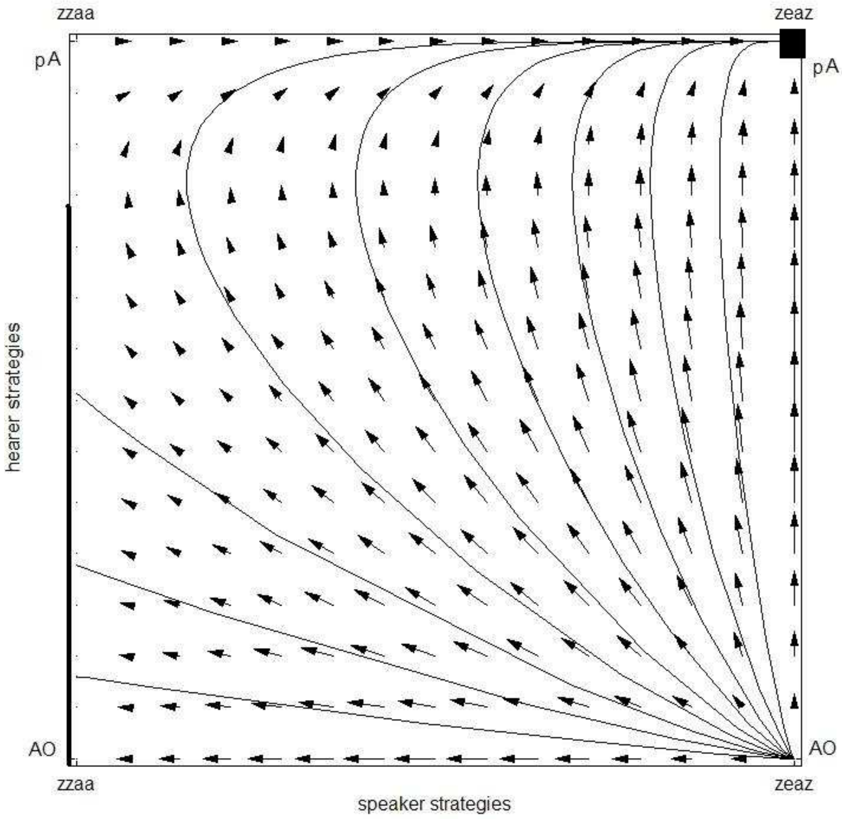

Figure 1: Dynamics without mutation

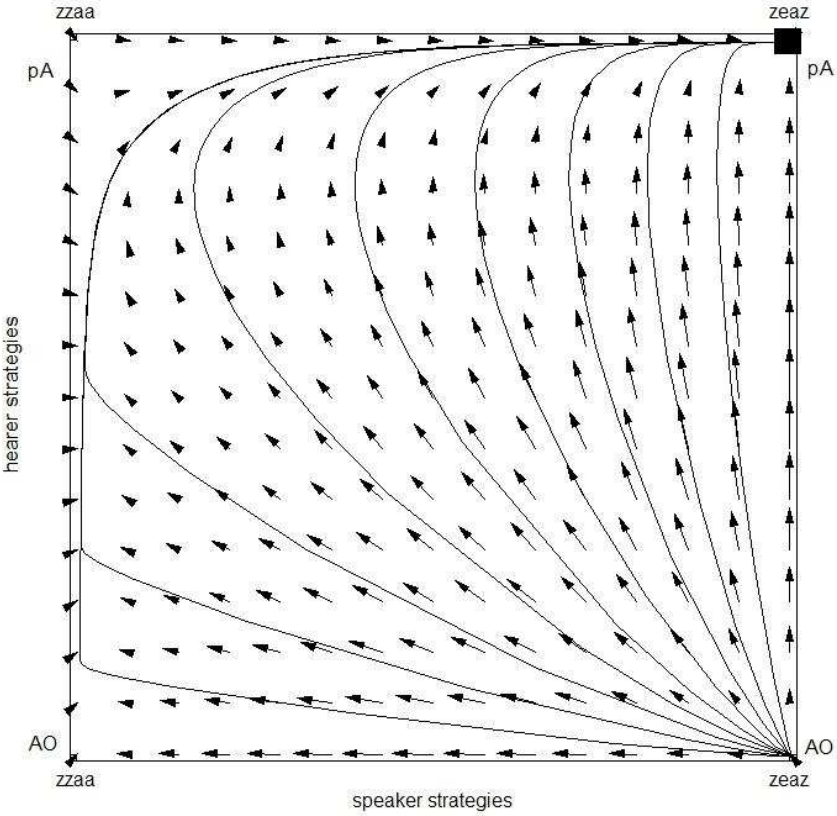

Figure 2: Dynamics with mutation

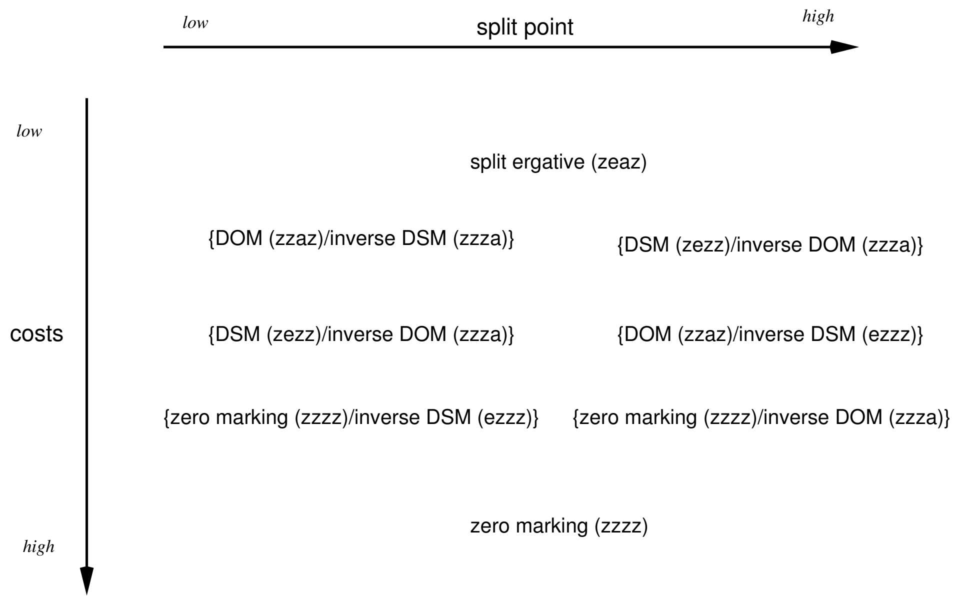

Figure 3: Configuration of Strict Nash Equilibria

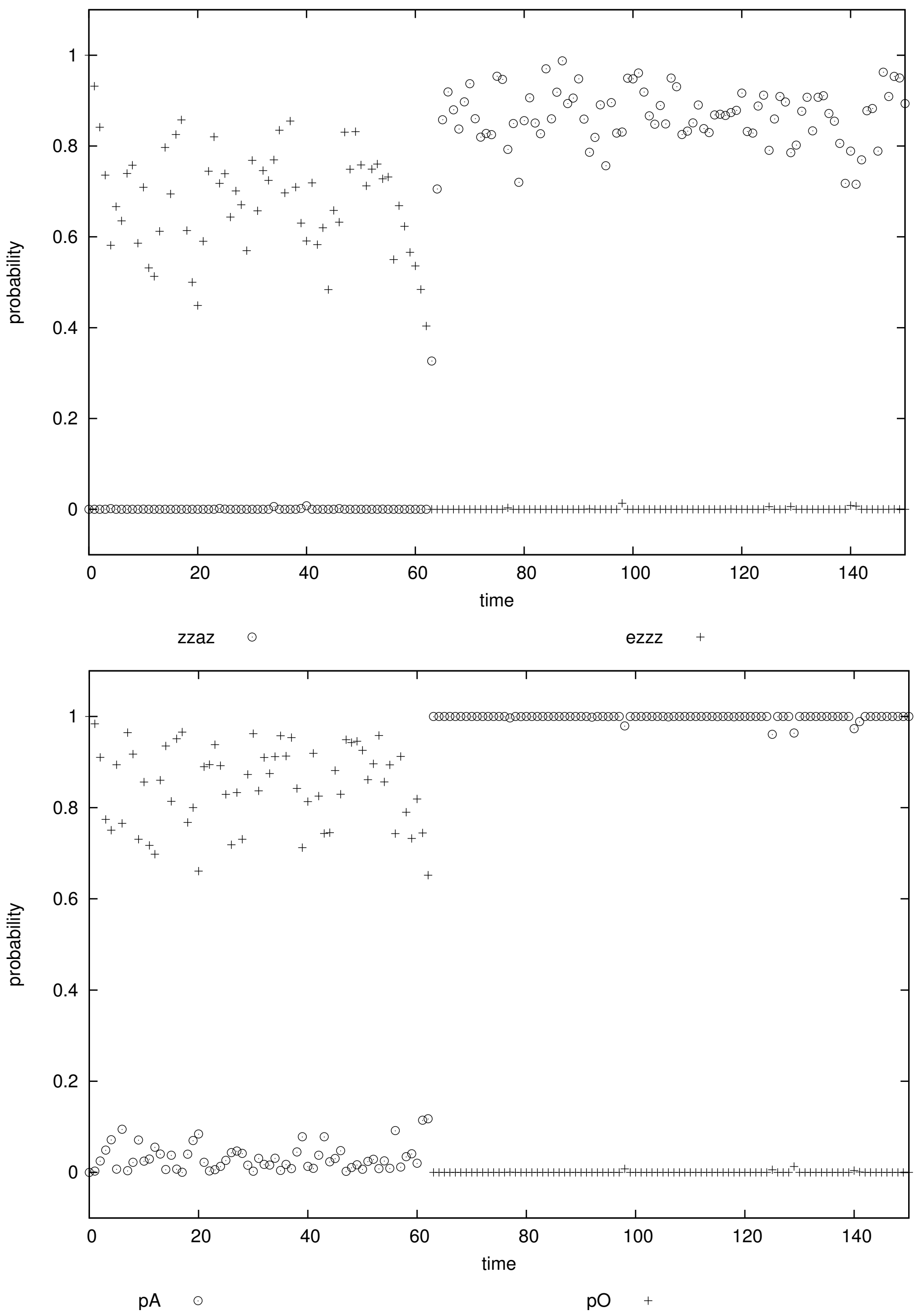

Figure 4: A simulation of stochastic evolution of the game from Table 8 above

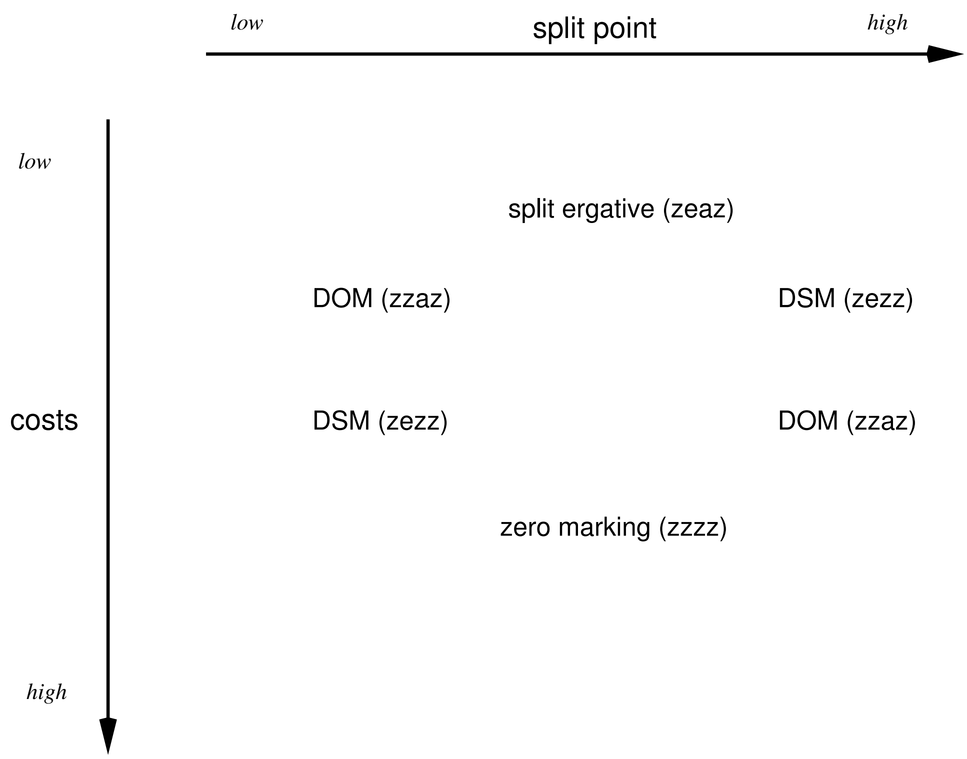

Figure 5: Configuration of stochastically stable states