

Drift, Convergence, and the Ergative Cycle

Paul Kiparsky
Stanford University

1 Three problems for the theory of language change

Over a century ago, Meillet (1918) identified as a major problem for historical linguistics the phenomenon he called *convergence* – independent parallel change in different languages, the linguistic counterpart of evolutionary biology’s *homoplasy*. A few years later, Sapir (1921) drew attention to the equally intriguing problem of *drift*, sustained unidirectional change by small increments. More recently many instances of drift and convergence have been recognized as parts of recurrent long-term *cyclic* trajectories (recently reviewed in Bouzouita et al. 2019; Gelderen 2023; Mosegaard Hansen & Waltereit 2025).

The evolution of Indo-Iranian case and agreement syntax instantiates all three challenges. Indo-Iranian inherited a Nominative/Accusative structural case system from its Indo-European ancestor. Its Indic and Iranian branches replaced it by an Ergative/Nominative system by independent but similar syntactic reanalyses. The languages in each branch then reverted towards a Nominative/Accusative system, step by step, via several different paths, in some cases passing through typologically unusual intermediate stages (Comrie 1978: 378, Stroński 2010). Some of the languages reveal the course of change by extensive dialectal variation in their case and agreement syntax, in others the loss of Ergative case has come to completion in unremarkable systems: Nominative/Accusative in Bangla and Farsi, no case at all in Central Kurdish. The entire two-phase process lasting two millennia is a classic instance of cyclic change. At least its second phase is also a case of drift and convergence. It is not unique to Indo-Iranian: some Pama-Nyungan languages of Australia are thought to have followed a similar trajectory (Bani & Klokeid 1974; McConvell 1981).

The three problems of drift, convergence, and cyclic change can be connected in an explanatory way by moving from *error-driven* models of change to a *bias-driven* model. Error-driven models, consistent with Saussure’s view of change as originating purely in *parole*, attribute language change to reanalysis by imperfect transmission. On this view change should be Brownian motion in typological space. In a bias-driven model, change is triggered and directed by learners’ prior hypotheses (“expectations” or “preferences”). It is an evolutionary process rooted in the principles that constrain language itself, in line with Jakobson (1929). But instead of the discredited theory of nomogenesis (or orthogenesis) that Jakobson appealed to, an OT-based bias-driven model of acquisition and change provides a non-teleological causal mechanism for the initiation, spread, and completion of linguistic innovations. It formalizes the biases by

violable constraints, which define what in biology has been called a *fitness landscape* (Gavrilets 2004), and convergent diachronic paths through it. In Kiparsky (2026) I presented theoretical arguments for the bias-driven model of change and its OT-based implementation, with supporting evidence from word-order change. It explains both why languages drift from SOV to SVO as they acquire an articulated spine of functional projections (Indo-European to English, for example), and why they revert from SVO to SOV as the syntactic heads become lexicalized as inflectional morphology (Austro-Asiatic to Munda, for example). This reconciles contradictory unidirectional drifts as phases of long-term cyclic change.

The present study applies the OT-based bias-driven model of diachronic change to the Indo-Aryan and Iranian dismantling of the Ergative/Nominative case system and the return to Nominative/Accusative morphosyntax. The development turns out to resemble that of word order in interesting ways. As with the change of SVO to SOV word order, the change of Ergative-Nominative to Nominative-Accusative morphosyntax was very likely caused by changes in the verb morphology, in the latter case specifically in the tense/aspect system (Jügel 2015: 441 ff., Haig 2017; Butt & Deo 2017). Another similarity between the changes is that both involve a bumpy ride from one fairly consistent canonical system to another via cross-linguistically more unusual transitional systems, such as the mix of ergative and nominative syntax in Nepali – still within the typological ballpark, but formally more complex in a precise sense to be defined below. Its analog in the word order domain is the mix of head-finality and head-initiality in older Germanic (and still in modern German) which is reached en route from Indo-European consistent head-finality to English and Scandinavian consistent head-initiality.

After outlining the theoretical assumptions and their consequences in section 2, I present the basic syntactic developments to be accounted for in section 3. The Middle Indic starting point is uncontroversially a system where subjects of transitive perfective verbs bear ergative case, while objects and intransitive subjects are nominative, and agree in gender and number with the verb. The case/agreement systems after this initial stage up to and including the modern languages show a broad range of variation and diachronic paths (for Indo-Aryan Butt 2001; Butt & King 2004; Deo & Sharma 2006; Stroński 2010; Butt & Deo 2017; Dahl & Stroński 2016; Butt 2017; Patel-Grosz 2021, for Iranian Haig 2008, 2017; Gündoğdu 2022; Kim 2025). In section 4 I then formulate a set of constraints whose possible rankings generate a typology, which is a small superset of the attested systems. The changes can be modeled as sequences of optimizing constraint rerankings, in accord with the core hypothesis of bias-driven diachrony.

2 Diachronic assumptions

2.1 The OT-based approach

We assume that grammars are partially ranked constraint systems whose outputs define the grammatical utterances of a language. In such a framework, change is by definition a reranking of constraints. We make two provisional simplifying assumptions: all rankings are a priori equally probable, and reranking proceeds in minimal

steps, moving a language to a state that is immediately adjacent to it in typological space.

Constraints constitute prior knowledge that enable language acquisition. This does not necessarily mean that they are innate. They might evolve in part through maturational processes. In the course of acquisition a basic constraint inventory could be systematically expanded into constraint families and stringency hierarchies. Moreover, since Jakobson (1941) babbling and linguistic play have been understood as pre-linguistic explorations of typological space.

The core idea is that learners are biased in favor of the most *probable* language consistent with the Primary Linguistic Data they have processed and internalized.

Optimality Theory provides an exact, simple, and testable definition of the probability of a language (Riggle 2009). The r -volume of language l under k constraints is the number of constraint rankings (grammars) that generate l , divided by $k!$. The probability of a language is equal to its r -volume. Canonical (“unmarked”) systems are r -volume maxima. This yields a conceptually attractive metric of complexity because it relies on the same ranking principles and constraints that characterize grammars and typological space. It is also a promising metric because it has already proved to be a good predictor of typological distribution (Bane & Riggle 2010) and of the relative frequency of free variants (Kiparsky 1993; Anttila 1997 *et seq.*), including syntax (Bane & Riggle 2010). It in effect defines formal markedness in a way that links typology with diachrony. In the case of word order, a remarkably simple OT constraint system derives the basic typology and an improved version of the Final-over-Final Constraint, and underpins the major diachronic pathways of innovations within that typological envelope (Kiparsky 2026).

The proposal is consistent with several plausible models for the way a learner starts from an initial state and ends with a grammar that outputs the target language. As the learner’s starting point we can either suppose a set of unranked constraints, or some fixed ranking. For the learning procedure we can either suppose that the learner generates random hypotheses consistent with the data, or that she at each stage of acquisition projects the language with the largest r -volume. The latter learner would obviously learn much faster, but at the price of a large computational overhead. Without committing to any of these alternatives, our abstract model predicts that the rate and time of acquisition of a subsystem or of an entire language will be proportional to its r -volume. At the same time, in concert with a bias-driven theory of change it derives interesting predictions about linguistic stability and the direction of language change, which in the cases studied so far have proved surprisingly accurate.

To test the hypothesis on a set of historically related grammars, we construct OT analyses for the subsystem of interest, pool the constraints that are active in each of them, and generate the factorial typology. Bias-driven acquisition then predicts that increasing r -volumes of language types define diachronic pathways. For tiny constraint systems the typology and ranking volumes can be calculated by hand. Larger constraint systems, such as the one investigated in this article, require computational tools, such as PyPhon (Bane et al. 2011).

2.2 Acquisition and evaluation metrics

The idea of bias-driven acquisition and change is not entirely new. The biases are comparable to the “attractor states” posited in recent typological and historical work (Haig 2008; Seržant & Moroz 2022; Haider 2023; Gelderen 2023), and to parameter settings (Dresher 1999). Unlike attractor states, they are not specified as grammatical states and configurations, but emerge intrinsically from elementary violable constraints whose rankings characterize a structured landscape of language states. In other words, the biases are not holistic prototypical ideal grammars or features of grammars, but emerge from the building blocks out of which grammar itself is constructed. This compositionality makes OT constraints conceptually more akin to unmarked parameter values than to attractor states. But unlike parameters, constraints are not mutually independent dimensions in typological space, but interact in accord with the principle of constraint domination to define grammars and typologies.

If the biases that drive change reflect OT constraints that the learner brings to the acquisition process, then several things follow. Generalizations about language change and the OT constraint inventory required by grammatical theory, typology, and language acquisition must be mutually consistent. Constraints cannot be stipulated just for the sake of diachrony. Furthermore, we know from typology that constraints are not just about purely formal structure, but have substantive content.

The acquisition literature offers two seemingly incompatible hypotheses about formal learning biases: that learners seek the *simplest* grammar consistent with the data encountered, and that they seek the *most restrictive* grammar. The former hypothesis, known as the simplicity criterion, is based on the idea that simplicity corresponds to generality, so that the search for the simplest grammar leads to the most general grammar that is consistent with their innate UG endowment. This has been a standard assumption since the earliest days of generative grammar (Halle 1962).

The contrary hypothesis that learners select the most restrictive grammar consistent with the data is based on the idea that a conservative learning strategy pays off because undergeneration can be corrected simply by positive evidence, whereas cutting back on overgeneration involves discovering the unacceptability of some structures – a harder task because it requires negative evidence from systematic gaps or explicit corrections. A formal articulation of this idea is the Subset Principle (Berwick 1985; Wexler & Manzini 1987, Clark & Roberts 1993: 304-5), according to which the learner at each stage of the learning procedure seeks the most restricted grammar that is compatible with the input, or more precisely with the intake, the part of the input that is actually processed and internalized. It correctly predicts the frequent pattern of unidirectional historical development towards narrower binding domains, in which pronominals become anaphors, long-distance anaphors become local anaphors, and so on (Kiparsky 2002, 2012b). But it does not extend to the well-documented class of changes that move towards the most general grammar, rather than to the most restricted one.

Our hypothesis reconciles these two opposite hypotheses, and explains how the preferred hypothesis depends on the nature of the constraint. In particular, it explains why the Subset Principle works in Binding Theory, whereas generality is key in case assignment. The explanation is briefly as follows.

Under the standard assumption that cross-linguistic variation in binding domains is governed by a set of constraints that impose disjoint reference on anaphoric elements, or require coreference on them, in a hierarchy of domains (such as the minimal clause, the finite domain, and the sentence), the simplest grammar is the one in which *all* these constraints take effect, which necessarily singles out the narrowest binding domain.

The opposite directionality is predicted for case assignment. In Iranian and Indo-Aryan for example, the case pattern of the present tenses was systematically generalized, in stepwise fashion and by different routes, to the past tenses: “no language changed its present tenses; no language took its past tenses further away from its present tenses” (Haig 2008: 325). The formal analysis presented below directly explains why it was the nominative-accusative case system that prevailed. The conflict between two verb agreement constraints, one requiring nominative agreement, the other requiring subject agreement, makes systems where subjects are ergative and objects are nominative inherently unstable in languages with verb agreement, because only one of these constraints can ever be satisfied in transitive clauses. The theory cashes out this conflict formally as a reduction of *r*-folume. Ergative subjects in languages without verb agreement present no such conflict, so that, other things being equal, ergative case should be more stable in them.

In sum, while pronouns start out free and acquire binding domains that become ever narrower in time, ending up as locally bound reflexives, case patterns are generalized from restricted domains to larger ones. From our perspective, there is no conflict here between two apparently incompatible formal principles – simplest grammar versus most restrictive grammar – for both changes follow a consistently optimizing path.

In contrast, optimization with respect to functionalist constraints on the mental computational system appears to be orthogonal to these formal asymmetries. They include the idea that grammars (and linguistic changes) favor the shortest derivation (which would minimize computation and speaker effort), shortest moves (minimizing long-distance dependencies to reduce burden on short-term memory), and MERGE over MOVE (shorter outputs), as proposed by Chomsky 1995 and in later minimalist work. Other functional criteria, defined specifically for language acquisition and with important implications for language change, would be Fitness (Yang 2000), which promotes the grammar that generates the highest frequency of unambiguous outputs, the Tolerance Principle (Yang 2016), and Hawkins’ (Hawkins 1994, 2004) Minimize Forms principle. Insofar as these preferences are definable on grammars, they can be expressed as constraints, and their import for linguistic change can be empirically explored.

3 The diachrony of Indic and Iranian case and agreement systems

3.1 Forward to the past

Indic and Iranian had evolved an ergative/nominative case system where Past/Perfective verbs agreed with nominative objects, and much of the case morphology, including

accusative case, was lost. They then returned to a nominative/accusative case system, also with nominative subject agreement, broadly similar to that of the original Indo-Aryan languages. The languages moved independently forward to the past, one step at a time. The exit strategy consisted of three distinct and at least in part independent overt changes:

- (1) a. Rebuilding accusative case,
- b. spread of subject agreement,
- c. loss of ergative case marking.

In all languages we find that case is in the lead and agreement “catches up” – the *agreement lag* noticed by Deo & Sharma (2006): “We never find a system in which the verb agrees with a non-nominative object when a nominative subject is available [even in Gujarati!], but we do find agreement with non-nominative subjects when a nominative object is available.” This generalization has proved to be valid, even in the face of challenges from Marwari and Kutchi Gujarati data, as discussed in section 4 below.

(2) is a schematic summary of the offramp, starting from the original Middle Indic ergative/nominative system before the loss of accusative case, and ending with vanilla nominative/accusative and caseless languages. The leftmost column shows the *r*-volumes, which rise steeply throughout the trajectory. Section 4.3 below explains how they are computed and presents the full typology.

(2)	.012	Middle Indic		Erg	Nom _i	Agr _i			
	.036	Gujarati/Hindi	Erg	Acc _i	Agr _i	∅	Erg	Acc	Agr ₀
	.104	Nepali			Erg _i	Acc			Agr _i
	.292	Bengali			Nom _i	Acc			Agr _i

3.2 The origin of Indic and Iranian ergative case

The historical source of ergative morphology is controversial, and our analysis is not committed to any particular provenience. An old hypothesis traces it to a Sanskrit instrumental case ending which originally marked the passive agent, but a seductive alternative derives it from dative $-\bar{e}$, presumably directly continued in Gujarati and Assamese $-e$ (Genetti 1993; Mohanan 1994; Butt 2001; Butt & King 2004; Butt 2006a,b; Li 2007; Butt & Ahmed 2011). The new clitics, Nepali $=ne$ and Marathi, Punjabi and Hindi/Urdu $=le$, might then have arisen by refurbishing $-e$ with an onset in order to match the $=CV$ shape of the other clitics in those languages.¹

¹Suffixes can be upgraded to clitics when they are part of a paradigm whose other members are clitics (examples from English, Irish, Mainland Scandinavian, Estonian, Seto, Võru, Saami, Spanish, and Greek in Kiparsky 2012a). The new clitics $=ne$ and $=le$ emerged in languages in which the other case endings are consistently clitics of the form $=CV(C)$, such as Hindi/Urdu dative/accusative $=ko$, instrumental $=se$, inessive $=m\bar{e}$, adessive $=par$, Marathi dative/accusative/locative $=la$ (earlier $=si$), Nepali $=l\bar{a}i$, $=ko$, $=m\bar{a}$.

3.3 The rise of a new accusative

In an early phase of the path back to nominative/accusative case marking, perhaps as its first step, an accusative case ending is reintroduced, first for definite/human object arguments. Such a new ending emerged in most Indic and Iranian languages, independently and from etymologically disparate sources. The newly accusative-marked object at first continues to agree with the finite verb. The immediate outcome of this innovation is thus an unusual system where, if the subject is ergative and the object is accusative, the finite verb agrees with the accusative object. This is famously found in Gujarati (Mistry 2004), as well as in the Kuḍaḷi dialect of Marathi (Deo & Sharma 2006).

- (3) rāj-e sītā-ne pajāv-ī
 Raj.MASC.ERG Sita.FEM.ACC harass-PERF.FEM.SG
 ‘Raj harassed Sita’ (Gujarati)

Accusative object agreement is most naturally derived from nominative object agreement as a result of the rise of accusative object marking, for a change from subject agreement to object agreement would be unmotivated.

Hindi, Marathi, Pahari and some Northern Kurdish dialects deal with the rise of accusative objects in a different way: by introducing 3rd person singular default agreement:

- (4) a. sītā-ne rādhā-ko dekh-ā
 Sita.FEM.ERG Radha.FEM.ACC see-PERF.MASC.SG
 ‘Sita saw Radha’ (Hindi, Deo & Sharma 2006, Bhatt 2023)

In Gujarati, where accusative objects agree with the verb, 3rd person neuter singular default agreement appears where the object is dative, as the downstairs agent of the causative in (5b) (Joshi 2020):²

- (5) a. priyanka-e ram-ne dawdaivo
 Priyanka.F.SG.ERG ramM.SG.-ACC run.CAUS.PERFECTIVE.M.SG.
 ‘Priyanka made Ram run.’
 b. priyanka-e ram-ne dawdaivu
 Priyanka.F.SG.ERG ramM.SG.-DAT run.CAUS.PERFECTIVE.N.SG.
 ‘Priyanka made Ram run.’

3.4 Spread of subject agreement

In a further change, the agreement trigger is transferred from nominative objects to ergative subjects. This happened in Nepali (with the abovementioned new ergative ending =*le*, Masica 1993; Li 2007), Marheṭhi (Deo & Sharma 2006), Eastern Pahāḍī, Shina, Assamese (Verbeke & Clercq 2016) and some Northern Kurdish dialects, even where the object remains nominative:

²As noted by Butt & Ahmed (2011: 555-6), the case variation is analogous to the variation in Sanskrit causatives between accusative and instrumental case; the respective Sanskrit counterparts of (5a) and (5b) are *rāmaṃ dhāvayati* and *rāmeṇa dhāvayati*.

- (6) a. mai-le mero lugā dho-en
I-ERG my clothes.NOM wash-perf.1SG
'I washed my clothes' (Nepali, Deo & Sharma 2006)
- b. mai-lē pāp garē
me-ERG sin-NOM did-1SG
'I sinned' (Pahāḍī, Chatterji 1926: 969)
- c. wan tu dît-in
they-ERG.PL you.SG.NOM see.PAST-3PL
'They saw you' (N. Kurdish, Gündoğdu 2017)

3.5 Retreat of ergative case marking

A third step towards a nominative-accusative case system is the progressive loss of ergative case endings, beginning with the speech act participant (SAP) pronouns (1. and 2. person). In Punjabi, the 1.Sg. ergative pronoun *mē*, and in some varieties also the 2.Sg. *tū*, are morphologically unmarked (Butt 2017: 823, Dahl & Stroński 2016: 24). The 1.Pl. and 2.Pl. ergative pronouns are unmarked in Old Marathi and Warhaḍhi Brahmaṇi (Deo & Sharma 2006). Many languages extend this to all SAP pronouns.

This change leads to another fork in the path. Newly stripped of their ergative case endings, and morphologically identical with nominatives, the subject pronouns can either continue to function syntactically as ergative, remaining immune to verb agreement, or they can begin to function syntactically as nominatives and agree with the verb. The former, more conservative solution is more common. It is reported for Punjabi, standard Marathi (e.g. the Pune variety), Marheṭhi, Old Marathi, and Warhaḍhi Brahmaṇi (Deo & Sharma 2006), Bagri (Gusain 2000; Stroński 2009), and for West Iranian (Haig 2008). The following examples are cited from these sources:

- (7) a. mē lakṛi vaḍ-i
I.ERG wood.F.Sg.Nom cut-PAST.F.SG
'I (male or female) cut the wood' (Punjabi, Butt 2017: 823)
- b. tē á kiṭab koni pəḍi hε
you.ERG this book.FEM.SG NEG read.PPP.FEM.SG be.AUX.2SG
'You haven't read this book.' (Bagri, Gusain 2000: 62)
- c. mī sita=lā bagha-to
I.M.NOM Sita.F-ACC see-PRES.M.SG
'I see Sita.' (Marathi)
- d. mī ek chimṇī baghit-lī
I.M.ERG one sparrow.F.NOM see-PERF.F.SG
'I saw a sparrow.' (Marathi)
- e. mī sita-lā baghit-la
I.M.ERG Sita.F-ACC see-PERF.N.SG
'I saw Sita.' (Marathi)
- f. hā=chi yogu āmhī vivasvat-ā kathi-lā
this=emph knowledge.M.NOM we.ERG Vivasvat-DAT relate-PERF.M.SG
'It is this knowledge (that) we (royal) related to Vivasvat.'
(13th c. Marathi, *Jñāneśvarī* 4:16)

When the verb cannot agree either with the ergative subject or the accusative object, it defaults to neuter singular.

Some varieties of Marathi show a more advanced stage, in which overtly unaffixed subject pronouns agree in person, number, and gender with the verb, in effect functioning as nominatives:

- (8) *mī devā-jaṅval tu-jhyā-sāmne pāp kelo*
 I.M.NOM God-near you-GEN.OBL-before sin.N.NOM do-PERF.1.M.SG
 ‘I committed a sin near God and in front of you.’ (Gowari Marathi, Deo & Sharma 2006).

Unsurprisingly, this stage appears to be diachronically fragile. A more transparent relation of agreement to overt case marking is restored in one two ways. Modern Hindi has regularized the morphological expression of Ergative case by the ergative clitic =*ne*, which is added even to SAP pronouns.

- (9) *mai=ne chiḍiyā dekh-ī*
 I.M.ERG sparrow.F.NOM see-PERF.F.SG.
 ‘I saw a sparrow.’ (Hindi, Butt:2017a)

The more radical resolution of the opacity is a generalized Gowari-type restructuring of all bare ergatives as true nominatives, re-establishing the original state in which the verb agrees uniformly with the grammatical subject – the highest nominative – in perfective as well as imperfective verbs. This happens in Bangla, Oriya, and Bhojpuri (and possibly already in Māgadhi Apabhraṃśa, from which they are descended, Chatterji 1926: 971-972), also independently in Dharwari Marathi (Deo & Sharma 2006), and on the Iranian side in Farsi and Gilan Kurdish (Haig 2008: 254).

- (10) a. *tu şutur-ê biyaban-ê venda k’ir-î*
 2S:DIR camel-OBL desert-OBL losing did-2SG
 ‘You lost the camel in the desert’ (Gilan Kurdish)
- b. *āmī sita-ke dekh-lām*
 I.Nom Sita.F-ACC see-PERF.1SG
 ‘I saw Sita.’ (Bangla)

3.6 Periphrastic perfectives and the elimination of ergative case marking

The loss of ergative case marking goes to completion in Marwari and some dialects of Gujarati spoken in Northwestern India. The older literature describes Marwari as marking ergative case only in third person pronouns, and some masculine plural nouns, and only optionally at that (Magier 1983: 310 ff., Stroński 2009, Verbeke 2013: 212).

- (11) a. *vo ~ uṅ kāi kariyo*
 he ~ he-ERG what-ACC did
 ‘What did he do?’
- b. *mhāro beṭo kāi kariyo*
 My.∅ son.NOM what-ACC did
 ‘What did my son do?’

Contemporary Marwari has completely lost ergative case marking (Chandra & D’Alessandro 2022), as has the Kutchi dialect of Gujarati (Patel-Grosz 2021).

These dialects have developed a periphrastic perfective in which the auxiliary agrees in person and number with the now nominative subject, and the participle agrees in gender and number with the object, in a construction which Magier (1983) and Patel-Grosz (2021) liken to the Romance periphrastic perfect.

- (12) mhẽ sītā-ne dek^hī hū
 I Sita..FEM-ACC seePPP.FEM. be.1.SG.

‘I have seen Sita.’ (Marwari, Magier 1983: 322)

In the simple past the verb is a bare participle with the auxiliary omitted. It shows no person agreement, only gender and number agreement with the object.

In earlier forms of Western Hindi that retain ergative case, the periphrastic perfective shows complete object agreement when the subject is marked with ergative case. The participle then agrees with the object in gender and number and the auxiliary agrees with the object in person, as in (13).

- (13) tē-nē tin đəgəhē mē kal pē se bəʃʌjə ũ
 you-ERG three place I death from save.PPP.MASC.SG. be.PRES.1SG.

‘You have saved me from death on three occasions.’ (Braj Bhāṣā, Stroński 2009: 100, from Liperovskij (1988) (*non vidi*))

An early stage of the perfective periphrasis with ergative attrition is found as early as in Gandhari Prakrit Niya texts from the 3rd century AD (Jamison 2000, Butt & Deo 2017: 543). A first or second person subject is endingless, and in the past tense it agrees in person and number with the auxiliary that is attached to the perfect participle in *-t(a)*. The third person singular lacks an overt auxiliary, but has an extended variant in *ǰa*.³

- (14) Singular Plural
 1. dit-emi dita-ma
 2. dit-esi dita-tha
 3. dita(ǰa) dita-ṃti

Since the third person singular verb has no auxiliary, it shows no agreement. If the subject and object are both human, and the subject is singular, it bears ergative/instrumental case.

- (15) a. ime isa veyā vari-da-ma
 them here we prevent-PPP-1PL
 ‘Here we have prevented them.’ (Jamison 2000:72)
 b. priyavaǰa-ṣa suǰute-na uṭa di-ta dhamana-ye
 priyavaǰa-DAT/GEN suǰuta-GEN/DAT camel give-PPP break-in-for
 ‘Suǰuta gave priyavaǰa a camel to break in’ (Jamison 2000:74)

³The 3P null auxiliary is familiar from the Sanskrit periphrastic future in *-tā* (*dātā-smi*, *dātā-si*, *dātā* ‘I/you/(s)he will give’), and is cross-linguistically common.

The complete loss of ergative case marking is also reported for the early Hindi of the 15th century poet Kabir, where not only pronouns but also nouns have lost their ergative case endings, while transitive perfective verbs show the participial agreement pattern with the object in number and gender (Butt & Deo 2017: 541).

- (16) *dās kabir jatan-se oḍh-i*
 servant Kabir.ERG care-with wrap-PERF.F.SG
 ‘Your servant Kabir wore it with great care.’ (Early Hindi)

In the next section we present the nine-constraint system that is the basis of the grammatical analysis, the ten-language typology that it generates, and the paths towards increasing *r*-volume through it.

4 The grammars and their evolution

4.1 Case and argument licensing

To model the case and agreement systems and their evolution from Middle Indic and Iranian to their descendants we use a framework that posits a level of Abstract Case and a level of morphosyntactic case (Butt 2006a; Galbraith 2022).⁴ The four Abstract structural Cases are essentially equivalent to the grammatical relations A, S, O, D that have been widely used in descriptive practice and typological work since Dixon (1979). They are assigned to arguments based on their Thematic Roles. Theta-roles (“who did what to whom”) are mapped into Abstract Case features according to their hierarchical rank. Transitive subjects have underlying Ergative Case and objects have underlying Accusative Case.

- (17) Abstract Cases
 Accusative: [−HR] “I’m not a subject”
 Ergative: [−LR] “I’m not a direct object”
 Dative: [−HR, −LR] “I’m neither a subject nor a direct object”
 Nominative: [] “whatever” (the default case)

Ranked violable Faithfulness and Markedness constraints then match Abstract Cases to their morphosyntactic counterparts – ergative, nominative, accusative, and dative morphosyntactic case, to agreement relations, and in some languages to word order. The mapping between Abstract Case and morphosyntactic case is simplified by decomposing both into bundles of the same cross-classifying features [\pm H(ighest)R(ole)], [\pm L(owest)R(ole)]. A Faithfulness constraint provides one-to-one mapping as the default; for example, morphosyntactic ergative and accusative case respectively correspond to abstract Ergative Case and abstract Accusative Case. The default correspondences are defeasible by higher-ranked Markedness constraints.

The featural decomposition allows natural classes to be captured by a single feature specifications. At the level of morphosyntactic case, the feature [−LR] captures the dative/ergative syncretism observed by DeLancey (1981); Butt (2006b), and (Butt & Ahmed 2011: 4.2), and [−HR] captures the dative/accusative syncretism that is ubiquitous in Indic and elsewhere, e.g. Hindi *-ko*, Punjabi *-nū*, Nepali *-lai*, Gujarati

⁴A very different approach is put forward in Bányai (2018: 161-170).

-ne (Joshi 2020; Deo & Sharma 2006: 379). At the level of Abstract Case, [+LR] specifies the class of internal arguments (direct objects and unaccusative subjects⁵), and [+HR] specifies the class of subjects (transitive and intransitive). For example, Hindi /-ko/, lexically specified as [-HR], can morphosyntactically unify with Abstract Accusative [-LR, -HR] and with Abstract Dative [+LR, -HR], and when it does, it acquires those respective specifications. Suffixes like Sanskrit *-ta* and English *-ee* (*invitee, standee, devotee, absentee*) are [-HR].

All languages can be assumed to have the same thematic relations, and therefore the same Abstract Cases. Here we concentrate on their morphosyntactic expression as cases and agreement. Word order can also express Abstract Case. Although it is generally free in South Asian languages, it does play a role, as shown by the Freezing effect first observed for Hindi by Mohanan (1989). In an ordinary transitive clause, all permutations of the subject, object, and verb are grammatical, with no effect on grammatical relations. But with experiencer predicates, switching the order of arguments changes grammatical relations. The thematically highest argument with dative case is a subject in subject position, and an object in object position. This is shown by the usual subjecthood tests of control, disjoint reference, and anaphoric coreference. In the examples below, the boldfaced argument is the grammatical subject.

(18) Dative experiencer subject

- a. **anuu-ko_i** niinaa_j khelte_{i,*j} hue dikhii *control*
 Anu-DAT Nina.NOM play-NF be-NF appear.PERF
 ‘Anu saw Nina while (Anu) was playing’
- b. **anuu-ko_i** niinaa_j uskii_{*i,j} bastii-mē dikhii *disjoint reference*
 Anu-DAT Nina.NOM her.GEN neighborhood appear.PERF
 Anu saw Nina in her (Nina’s) neighborhood’
- c. **anuu-ko_i** niinaa_j aapnii_{i,*j} bastii-mē dikhii *anaphora*
 Anu-DAT Nina.NOM self.GEN neighborhood-LOC appear-PERF
 Anu saw Nina in her (Anu’s) neighborhood’

Without any change in the cases, switching the word order changes the subject from *Anu* to *Nina*:

(19) Dative experiencer object

- a. **niinaa_i** anuu-ko_j khelte_{i,*j} hue dikhii *control*
 Nina.NOM Anu.DAT play-NF be-NF appear.PERF
 ‘Anu saw Nina while (Nina) was playing’
- b. **niinaa_i** anuu-ko_j uskii_{*i,j} bastii-mē dikhii *disjoint reference*
 Nina.NOM Anu-DAT her.GEN neighborhood-LOC appear.PERF
 ‘Anu saw Nina in her (Anu’s) neighborhood’
- c. **niinaa_i** anuu-ko_j aapnii_{i,j} bastii-mē dikhii *anaphora*
 Nina.NOM Anu-DAT self.GEN neighborhood-LOC appear.PERF
 ‘Anu saw Nina in her (Anu’s/Nina’s) neighborhood’

⁵But not including unergative subjects, if these are analyzed as having a silent cognate object or some other lower silent Theta-role.

As in many languages, structural case in Indic and Iranian languages is sensitive to tense/aspect features, definiteness, agentivity, animacy, and topicality. Only the two most important factors are included in the present analysis. First, Abstract Ergative corresponds to morphosyntactic ergative in subjects of Past/Perfect clauses, but not in subjects of Present/Imperfective verbs. This does not seem to be simply a matter of semantics or of some flavor of little *v*. Rather, it is determined by the morphological stem. Transitive verbs whose stems are morphologically past but have non-past interpretations still require ergative subjects (Haig 2008; Patel-Grosz 2021: 10).

Secondly, Abstract Accusative Case can correspond to morphosyntactic accusative case and to morphosyntactic nominative case (DOM). Accusative case endings on nominals that have some combination of features variously reported to include animacy, human-ness, lexical category (noun vs. determiner), definiteness, specificity, information structure (primary and secondary topicality), and affectedness (Mohanani 1994; Mistry 1997; Butt & King 2003; Li 2007; Dalrymple & Nikolaeva 2011; Montaut 2018; Witzlack-Makarevich & Seržant 2018; Lahiri 2021; and, for Farsi, Jasbi 2020: 71). A broad generalization is that abstract Accusative tends to correspond to morphosyntactic accusative in nominals that are less likely to be objects (Aissen 2003). Once the details are sorted out, a formal typology of DOM can be built.⁶ For now, I will just cavalierly refer to the relevant combination of features as the DOM (Differential Object Marking) feature, and mark it by a subscript *d* on the relevant nominal.

4.2 The constraints

The constraint system comprises a Faithfulness (MAX) constraint for each Abstract Case feature, which requires it to be matched by the featurally identical morphosyntactic case, and a Markedness condition that prohibits that morphosyntactic case altogether. Ranked that way, they generate split case systems such as DOM.

- (20) a. MAX-ERG(Past): Abstract Ergative corresponds to morphosyntactic ergative in clauses headed by Past/Perfect verbs.
- b. *ERG: No ergative.
- c. MAX-DOM(Acc): Abstract Accusative corresponds to morphosyntactic accusative in DOM nominals.
- d. *ACC: No accusative.

A full grammar would have to specify the conditions under which overt accusative and ergative case must, may, and cannot appear in the language.

A kind of counterpart to DOM for Ergative case assignment privileges features at the opposite end of the spectrum. For example, in Marathi 1. and 2. person Ergative pronouns have no overt ergative endings, yet function morphosyntactically as ergative.

⁶According to Montaut, the accusative is “obligatory in Hindi only with human individuated objects, and optional with inanimate objects even when individuated.” According to Jasbi, Farsi (formal Teheran variety) has obligatory *rā* on proper nouns, personal and demonstrative pronouns, reflexive pronouns, reciprocal pronouns, demonstrative nouns, superlatives, question-words *which* and *who*, strong quantifiers such as *each*, *all*, *most*, *both*, and plurals marked with *hā*.

At least in the languages under study this skewed distribution can be treated purely within the morphology. Unaffixed pronouns that function morphosyntactically as ergatives can be assumed to receive ergative case in the morphology, either by default or by a null suffix. The constraint evaluation then starts with the morphosyntactic ergative case in place.

The details of cross-linguistic variation are in principle treatable in OT by constraint ranking, but there is not enough information to test our frequentistic predictions. Since the varieties of person and number splits do not appear to interact with the major case and agreement parameters, their inclusion would not add much to the analysis, while swelling the typology exponentially. For now, therefore, I just define the MAX constraints on Past/Perfect and the diacritic *d* (mnemonic for “DOM”, “definite”, or “Determiner”), leaving open the specific distribution of *d* in each language.

In the absence of an eligible agreeing subject, some languages allow verb agreement requirements to be satisfied by default. This is treated here as agreement with a silent expletive pronoun that has unmarked case, number, and gender but no meaning and no Theta-role. In Hindi/Urdu the verb then gets third person masculine singular (Mohanan 1994). In accord with the Richness of the Base principle, the input may have expletives inserted freely, but the constraint system guarantees that they will surface in the optimal output of a derivation only where agreement is required and no other trigger of it is available. No ranking of the constraints can generate obligatory expletive objects, or non-nominative expletive subjects. The reason is that expletives in this system are motivated only by the agreement requirement, and other things being equal the constraints always prefer subject agreement over object agreement. This amounts to a formal prediction of the system, so if it turns out to be incorrect, additional constraints would be required, with further empirical consequences.

Our constraints are broadly similar to those of Deo & Sharma (2006). However, we do not use scalar constraints, constraint conjunction, or harmonic alignment, only standard markedness and faithfulness constraints. Our constraints are of two types: undominated well-formedness constraints, which must be satisfied both in input and in output representations, and ranked violable constraints. (In our computational implementation they were part of GEN rather than EVAL). The inviolable well-formedness constraints define basic clause structure:

(21) Unviolable dominant constraints

- a. Nominals have case.
- b. A finite verb agrees with exactly one nominal (overt or expletive).

The action lies in the violable constraints, which are enumerated in (22).

(22) Ranked violable constraints

- A. *AGR(Erg): Ergative arguments don’t agree with the verb.
- B. *AGR(Acc): Accusative arguments don’t agree with the verb.
- C. MAX-PAST(Erg): In Past/Perfect clauses, Abstract Ergative corresponds to morphosyntactic ergative.
- D. MAX-DOM(Acc): In DOM nominals, Abstract Accusative corresponds to morphosyntactic accusative.

- E. *ACC: No accusative.
- F. *ERG: No ergative.
- G. AGR_{Nom}: The verb agrees with a nominative.
- H. AGR_{Hi}: The verb agrees with the [+HR] argument (the argument bearing the highest Theta-role).
- I. *∅: No expletive.

Each constraint plays an active role in the synchronic grammar of at least one of the languages.

In the theoretical literature ergative case tends to be treated as a special problem, or as somehow more “marked” than accusative case. This is arguably eurocentric, because about 40% of the languages of the world that have case at all have ergative case. The case theory proposed here normalizes ergative case. Ergative and accusative are formally parallel: each constraint that involves one of them has a identical counterpart that involves the other. The theory would be *more* complex if there were no such thing as ergative case. The normalization also means that “ergative languages” are just languages that have ergative case, and “ergativity” is the property of having it. As far as I can tell, these categories have no profound typological significance, any more than “dativity” and “dative languages”, or “ablativity” and “ablative languages” have.

The apparent asymmetry between accusative and ergative case is due to the fact that ergative case and nominative case compete with each other (but not with accusative case) for agreement, but not on a level playing field. This is because of a conflict between two requirements: agreement with the argument that bears the most prominent Theta-role ([+HR]), and agreement with nominative case (enforced by the prohibition of agreement with marked cases accusative, ergative, and dative). Both cannot be satisfied together when transitive subjects are ergative. A grammar must resolve this intrinsic contradiction by ranking, which decreases its *r*-volume, hence by hypothesis decreases its learnability. A language with nominative subjects does not have this complication. The diachronic prediction follows that ergative case and accusative case are equally stable in a language without verb agreement, but ergative case is more vulnerable to loss in a language with verb agreement.

A further caveat before we proceed to the typology. While the constraints are basic and symmetrical, and omitting any of them would lose empirical coverage, some of them could be reformulated to generate the same set of outputs. In particular, the contextual conditions involving Tense/Aspect and DOM can be moved from the Faithfulness constraints to the corresponding antagonistic Markedness constraints. Instead of MAX-ERG(Past), MAX-ACC(DOM), *ERG, *ACC as above, we would then have across-the-board MAX-ERG, MAX-ACC, and contextually restricted *ERG(Non-Past), *ACC(Non-DOM). I do not know whether there are theoretical and empirical grounds for choosing between these and perhaps other alternative formulations of the constraints.

4.3 *r*-volumes and typological space

With this constraint system in hand, we compute the typological space and the ranking volume for the relevant subsystem of each language, which is the number of

grammars that generate it. The constraints generate a total of $9! = 362,880$ grammars, which converge on only ten possible languages, with a wide spread of ranking volumes.

(23) compactly displays five selected input structures and their respective outputs in each of the ten grammars generated by the totality of constraint rankings. They are listed in order of increasing ranking volume (the proposed predictor of frequency/-probability/structural simplicity), which is shown in the rightmost column. E, A, and N stand for Ergative, Accusative, and Nominative – as Abstract Case in the input, and as their morphosyntactic correspondents in the ten output languages. Nominals are marked with the subscript _d (which is specified in the input) to indicate that they bear the DOM feature, and with the subscript _i (which is derived in the output) marking the nominal that agrees with the finite verb of the clause. The first column shows the structure of Nonpast/Imperfective (I) clauses, in which morphosyntactic ergative case is barred, in which subjects always have nominative case, and in which the verb always agrees with the first or only nominative. The only variation in them involves the morphosyntactic accusative (A) in the case repertoire, absent in two forms of Middle Indo-Aryan and again in Central Kurdish, and present elsewhere. The remaining columns show Past/Perfective (P) clauses, where the intersection of variation in the presence or absence of ergative case, the DOM effect on accusative case, and the different types of agreement fills most of the typological space. Three of the ten output language types are to my knowledge unattested.

(23) Skeletal syntactic typology

Input	/IEA _d V/	/PEAV/	/PEA _d V/	/PE _d AV/	/PE _d A _d V/	<i>r</i> -volume
Undocumented	[IN _i A _d V]	[PEN _i V]	[PN _i A _d V]	[PE _d N _i V]	[PN _{di} A _d V]	.007
MIA	[IN _i A _d V]	[PEN _i V]	[PEN _{di} V]	[PE _d N _i V]	[PE _d N _{di} V]	.012
Undocumented	[IN _i A _d V]	[PEN _i V]	[PE _i A _d V]	[PE _d N _i V]	[PE _{di} A _d V]	.014
Hindi	[IN _i A _d V]	[PEN _i V]	[P _{0i} EA _d V]	[PE _d N _i V]	[P _{0i} E _d A _d V]	.036
Gujarati	[IN _i A _d V]	[PEN _i V]	[PEA _{di} V]	[PE _d N _i V]	[PE _d A _{di} V]	.036
MIA	[IN _i N _d V]	[PEN _i V]	[PEN _{di} V]	[PE _d N _i V]	[PE _d N _{di} V]	.104
(Gandhari)	[IN _i N _d V]	[PE _i NV]	[PE _i N _d V]	[PE _{di} NV]	[PE _{di} N _d V]	.104
Nepali	[IN _i A _d V]	[PE _i NV]	[PE _i A _d V]	[PE _{di} NV]	[PE _{di} A _d V]	.104
Bangla	[IN _i A _d V]	[PN _i NV]	[PN _i A _d V]	[PN _{di} NV]	[PN _{di} A _d V]	.292
C.Kurdish	[IN _i N _d V]	[PN _i NV]	[PN _i N _d V]	[PN _{di} NV]	[PN _{di} N _d V]	.292

To illustrate the derivations in detail, I present three sample tableaux, followed by summary descriptions of all ten languages that unpack the condensed information in (23).

In Bangla and Farsi, the subject is Nominative, the object is Nominative or (if it bears the DOM feature) Accusative, and the verb agrees with the subject. This very simple system (*r*-volume .292) is generated by 105,840 constraint rankings, one of which is shown in tableau (24):

(24) Bangla

		A	B	C	D	E	F	G	H	I
		*ERG	AGR _{Nom}	MAX-ACC(DOM)	*AGR _{Erg}	*AGR _{Acc}	AGR _{Hi}	*∅	MAX-ERG(Past)	*ACC
	Erg _d Acc _d V _{Perf}									
1.	Erg _i Acc Agr	*			*					*
2.	Erg Acc _i Agr	*				*	*			*
3.	∅ _i Erg Acc Agr	*						*		*
4.	Erg _i Nom Agr	*	*	*	*					
5.	Erg Nom _i Agr	*		*			*			
6.	∅ _i Erg Nom Agr	*		*				*		
⇒ 7.	Nom _i Acc Agr								*	*
8.	Nom Acc _i Agr		*			*	*		*	*
9.	∅ _i Nom Acc Agr		*					*	*	*
10.	Nom _i Nom Agr			*					*	
11.	Nom Nom _i Agr		*	*			*		*	

Promotion of *ACC in (24) yields a caseless system, such as that of Central Kurdish (same *r*-volume of .292).

In the more intricate system of Hindi/Urdu, the Agentive subject of a perfective transitive clause (Erg_{Ag}) is Ergative, a definite or animate object (Acc_d) is Accusative, and in such configurations the verb (V_{Perf}) has default 3Sg. Masc. Person/Number agreement, treated here as agreement with a null expletive. One of the 12,951 rankings that output Hindi/Urdu is (25):

(25) Hindi/Urdu

		A	B	C	D	E	F	G	H	I
		*AGR _{Erg}	*AGR _{Acc}	MAX-ERG(Past)	MAX-ACC(DOM)	*ACC	*ERG	AGR _{Nom}	AGR _{Hi}	*∅
	Erg _{Ag} Acc _d V _{Perf}									
1.	Erg _i Acc Agr	*				*	*			
2.	Erg Acc _i Agr		*			*	*		*	
⇒ 3.	∅ _i Erg Acc Agr					*	*			*
4.	Erg _i Nom Agr	*			*		*	*		
5.	Erg Nom _i Agr				*		*		*	
6.	∅ _i Erg Nom Agr				*		*			*
7.	Nom _i Acc Agr			*		*				
8.	Nom Acc _i Agr		*	*		*		*	*	
9.	∅ _i Nom Acc Agr			*		*		*		*
10.	Nom _i Nom Agr			*	*					
11.	Nom Nom _i Agr			*	*			*	*	

In Nepali, subjects of perfective transitive clauses, and inanimate subjects even in intransitive clauses, are obligatorily ergative.⁷ Animate subjects of imperfective transitives, and subjects of ergative intransitive perfectives, are optionally ergative, with a great deal of interspeaker variation (Li 2007, Verbeke 2013: 4.2, Lindemann 2024). “High-animacy” objects are accusative.⁸ Verbs agree with their subjects, irrespective of case-marking. 37,800 constraint rankings yield this output (modulo the above-mentioned idealizations concerning the case marking variation), one of which is (26):

⁷The undominated constraint that bars ergative case thus ends up applying only to animate subjects in Nepali.

⁸Proper names and pronouns obligatorily, nouns with specific reference optionally.

(26) Nepali

		A	B	C	D	E	F	G	H	I
		*AGR _{Acc}	MAX-ERG(Past)	MAX-ACC(DOM)	*ACC	*ERG	AGR _{Nom}	AGR _{Hi}	*∅	*AGR _{Erg}
	Erg _d Acc _d V _{Perf}									
⇒ 1.	Erg _i Acc Agr				*	*				*
2.	Erg Acc _i Agr	*			*	*		*		
3.	∅ _i Erg Acc Agr				*	*			*	
4.	Erg _i Nom Agr			*		*	*			*
5.	Erg Nom _i Agr			*		*		*		
6.	∅ _i Erg Nom Agr			*		*			*	
7.	Nom _i Acc Agr		*		*					
8.	Nom Acc _i Agr	*	*		*		*	*		
9.	∅ _i Nom Acc Agr		*		*		*		*	
10.	Nom _i Nom Agr		*	*						
11.	Nom Nom _i Agr		*	*			*	*		

Recall that the subscript _d marks an argument that is sufficiently high on the scale of Definiteness, "Animacy", Topicality, etc. to trigger morphosyntactic accusative marking when it is on Abstract Accusative argument, by constraint MAX-ACC(DOM).

4.4 The typology

The constraints generate altogether ten languages, described here with thumbnail sketches, again in order of ascending ranking volume.

- (27) The least probable grammar that is predicted to be possible, unattested
- Ergative, nominative, and accusative.
 - Every sentence has a nominative argument.
 - A finite verb agrees with the most prominent nominative argument.
 - Imperfective: nominative subject, accusative object.
Perfective: ergative subject if object is nominative, else nominative subject.
 - Ranking Volume: $2376/9! = 0.006548$
- (28) Middle Indic and Middle Iranian before the loss of accusative case
- Ergative, nominative and accusative.
 - Every sentence has a nominative argument.
 - A verb agrees with the most prominent nominative argument.
 - Perfective: in transitive clauses, ergative subject and nominative object, in intransitive clauses, nominative subject.
Imperfective: nominative subject, accusative object.
 - Ranking Volume: $4446/9! = 0.012252$.
- (29) A low-frequency system, unattested but predicted to be possible
- Ergative, nominative, and accusative.

- b. A finite verb agrees with the most prominent nominative argument.
- c. If there is no nominative, a finite verb agrees with the (Ergative) subject.
- d. Perfective: in transitive clauses, ergative subject, accusative and nominative object (DOM), in intransitive clauses, nominative subject.
Imperfective: nominative subject, accusative object.
- e. Ranking Volume: $5076/9! = 0.013988$

(30) Hindi/Urdu

- a. Ergative, nominative, and accusative.
- b. A finite verb agrees with the most prominent nominative argument.
- c. If there is no nominative, the verb agrees with a (3Sg.Masc) expletive subject.
- d. Perfective: ergative subject, accusative and nominative objects (DOM), in intransitive clauses, nominative subject.
Imperfective: nominative subject, accusative object.
- e. Ranking Volume: $12951/9! = 0.035689$

(31) Gujarati (after the rise of DOM)

- a. Ergative, nominative, and accusative.
- b. A finite verb agrees with the most prominent nominative.
- c. If there is no nominative, a finite verb agrees with the accusative object.
- d. Perfective: ergative subject, accusative and nominative objects (DOM).
Imperfective: nominative subject, accusative and nominative objects (DOM).
- e. Ranking Volume: $12951/9! = 0.035689$

(32) Middle Indic and Middle Iranian after the loss of accusative case, Balochi

- a. Ergative and nominative.
- b. Every sentence has a nominative argument.
- c. A finite verb agrees with the most prominent nominative.
- d. Perfective: ergative subject, nominative object.
Imperfective: nominative subject, nominative object.
- e. Ranking Volume: $37800/9! = 0.104167$

(33) Gandhari-type, but with restricted ergative marking and agreement, precursor of Nepali?

- a. Ergative and nominative.
- b. Every sentence has a nominative argument.
- c. A finite verb agrees with the most prominent argument.

- d. Perfective: ergative subject, nominative object.
Imperfective: nominative subject, nominative object.
- e. Ranking Volume: $37800/9! = 0.104167$

(34) Nepali, Assamese, Shina

- a. Ergative, nominative, accusative.
- b. A finite verb agrees with the most prominent argument.
- c. Perfective verbs have ergative subjects, imperfective verbs have nominative subjects, though in Nepali ergative is optional in most imperfective contexts.
- d. Objects are accusative and nominative (DOM).
- e. Ranking Volume: $37800/9! = 0.104167$

(35) Bangla, Farsi

- a. Nominative, accusative.
- b. A finite verb agrees with the subject (the highest argument = the highest nominative).
- c. Perfective and imperfective: Nominative subject, accusative and nominative objects (DOM).
- d. Ranking Volume: $105840/9! = 0.291667$

(36) Central Kurdish

- a. No case
- b. A finite verb agrees with the subject (the highest argument).
- c. Ranking Volume: $105840/9! = 0.291667$

As noted at (2), the unattested systems have low r -volumes, and the projected historical progression from shows a steady increase in r -volume. Interestingly, the Middle Indic loss of accusative case was a substantial simplification, increasing the r -volume from .012 to .104. This suggests that the dialectal differentiation of the Indic languages started *before* the loss of accusative case.

4.5 Generalizations

Certain systems cannot be derived by any ranking of these constraints. These are predicted gaps in the typology. The most important generalization is that there is no language in which subjects are nominative, but finite verb agreement targets objects (whether nominative, accusative, or some other case).

- (37) a. *Nom Acc_i Agr_i
- b. *Nom Nom_i Agr_i

To my knowledge, no such languages have been found in the Indo-Iranian family, nor anywhere else for that matter. Kutchi Gujarati and Marwari are not exceptions, since their perfective verbs are periphrastic, consisting of a finite auxiliary and a participle. The auxiliary (ellipsed in the simple past) agrees in person with the nominative subject, and the participle that agrees with the object in gender and number is not a finite verb, but a nominal form.

Generalization (37a), that a verb never agrees with a non-nominative object when there is a nominative subject (though it can agree with a non-nominative subject when there is a nominative object) has been frequently formulated in various ways (Anderson 1976, 1977, 1984; Comrie 1978; Dixon 1979, 1994; Bobaljik 2008), but it has never been formally explained. Here it follows, in a generalized form that also covers (37b), as a consequence of the constraints AGR_{Nom} and AGR_{Hi} , which also play a crucial role in deriving the diachrony.

A diachronic corollary is that nominative subjects never spread first: $Erg\ Nom_i$ $Agr_i \not\rightarrow Nom\ Nom_i\ Agr_i$. Our analysis predicts this, for the constraint system cannot derive agreement pattern (37b): it is *harmonically bounded*, as we have just seen.

The question raised by the historical trajectories is why the ergative/nominative case marking pattern of the perfective/past gives way to the nominative/accusative case marking pattern of the imperfective/present rather than the other way round. One possible answer is that ergative/nominative syntax is somehow unmarked in comparison to nominative/accusative syntax. This is dubious, because ca. 40% of languages that have case at all have an ergative. Another possible answer is that unmarked categories are the privileged basis for formal generalization (Kuryłowicz's 4th Law of Analogy), which in this case in this case means that the imperfective case, which is unmarked, is transferred to the perfective. Our analysis offers a third answer: *nominative/accusative syntax is optimal in a language with verb agreement*. This is because *all* agreement constraints are satisfied by canonical transitive clauses in languages with nominative/accusative case systems, but at most one of them is satisfied by canonical transitive clauses in languages with ergative/nominative case systems. Other things being equal, a language with verb agreement therefore has a necessarily greater *r*-volume if it has a nominative/accusative case system than if it has an ergative/nominative case system, regardless of whether the ergative is split or across the board.

To summarize the theoretical points: change is an evolutionary process governed by the same principles as language itself. Acquisition and change are guided by learners' prior expectations ("biases"). Drift moves towards the grammars privileged by those expectations. Lexical/morphosyntactic change may trigger radical reanalysis that restores the original type. This results in long-term diachronic cycles, in which the stepwise nature of change forces mixed, suboptimal systems to appear at intermediate stages in at least one phase of a cycle.

If our view of change is on the right track, it removes a worrisome paradox. Unidirectional trajectories of change are well documented. Yet historical linguistics reveals no decrease in typological variation at least in the last five millennia or so. How can irreversible changes be reconciled with the overall stability of typological diversity? The problem goes away if apparently unidirectional changes are parts of larger cyclic trajectories. Then we can confidently assume that no structural feature of human language has been irrevocably lost by language change, and any feature that appears

in the future is likely to have already existed in the past. Proto-languages were not qualitatively different from existing languages. This is good news for comparative historical grammar, for it means that we can safely continue to use typological generalizations as guidelines and guardrails in linguistic reconstruction.

Acknowledgements

[To be added]

References

- Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21(3):435–483.
- Anderson, S. R. 1976. On the notion of subject in ergative languages. In Charles Li (ed.), *Subject and topic*, 1–23. New York: Academic Press.
- Anderson, Stephen R. 1977. On the mechanisms by which languages become ergative. In Charles Li (ed.), *Mechanisms of syntactic change*, 317–363. Austin: University of Texas Press.
- Anderson, Stephen R. 1984. On representations in morphology: Case, agreement, and inversion in Georgian. *Natural Language and Linguistic Theory* 2(2):157–218.
- Anttila, Arto. 1997. *Variation in Finnish phonology and morphology*: Stanford University dissertation.
- Bane, Max, Sam Bowman & Jason Riggle. 2011. pyphon 1.5. Software package. <https://code.google.com/archive/p/clml/>.
- Bane, Max & Jason Riggle. 2010. The typological consequences of weighted constraints. *CLS Proceedings* 45.
- Bani, E. & T. J. Klokeid. 1974. Ergative switching in Kala Lagau Langgus. In P. Sutton (ed.), *Languages of Cape York: Papers presented to the linguistic symposium*, Australian Institute of Aboriginal Studies. <https://hdl.handle.net/11858/00-001M-0000-0012-74E0-7>.
- Bárányi, András. 2018. *Person, case, and agreement*. Oxford: Oxford University Press.
- Berwick, Robert C. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.
- Bhatt, Rajesh. 2023. February 21, 2003 Discussion of Deo and Sharma (2002). <https://web.mit.edu/rbhatt/www/24.956/14.pdf>.
- Bobaljik, Jonathan David. 2008. Where’s phi? Agreement as a postsyntactic operation. In Daniel Harbour, David Adger & Susana Béjar (eds.), *Phi theory: Phi-features across modules and interfaces*, 295–328. Oxford: Oxford University Press.

- Bouzouita, Miriam, Anne Breitbarth, Lieven Danckaert & Elisabeth Witzgenhausen. 2019. *Cycles in language change*. Oxford: Oxford University Press.
- Butt, Miriam. 2001. A re-examination of the accusative to ergative shift in Indo-Aryan. In Miriam Butt & Tracy Holloway King (eds.), *Time over matter*, Stanford, CA: CSLI.
- Butt, Miriam. 2006a. The dative-ergative connection. In O. Bonami & P. Cabredo Hofherr (eds.), *Empirical issues in syntax and semantics*, vol. 6, <http://www.cssp.cnrs.fr/eiss6>.
- Butt, Miriam. 2006b. *Theories of case*. Cambridge: Cambridge University Press.
- Butt, Miriam. 2017. Hindi/Urdu and related languages. In Jessica Coon, Diane Massam & Lisa Demena Travis (eds.), *The Oxford handbook of ergativity*, 807–831. Oxford: Oxford University Press.
- Butt, Miriam & Tafseer Ahmed. 2011. The redevelopment of Indo-Aryan case systems from a lexical semantic perspective. *Morphology* 21(3):545–572.
- Butt, Miriam & Ashwini Deo. 2017. Developments into and out of ergativity: Indo-Aryan diachrony. In Jessica Coon, Diane Massam & Lisa Demena Travis (eds.), *The Oxford handbook of ergativity*, 501–529. Oxford: Oxford University Press.
- Butt, Miriam & Tracy Holloway King. 2003. Case systems: beyond structural distinctions. In Ellen Brandner & Heike Zinmeister (eds.), *New perspectives on case theory*, 53–82. Stanford, CA: CSLI.
- Butt, Miriam & Tracy Holloway King. 2004. The status of case. In Veneeta Dayal & Anoop Mahajan (eds.), *Clause structure in South Asian languages*, 153–198. Dordrecht: Springer Netherlands.
- Chandra, Pritha & Roberta D’Alessandro. 2022. Ergativity, agreement and alignment shift in Western Indo-Aryan. <https://doi.org/10.21203/rs.3.rs-2288028/v1>.
- Chatterji, Suniti Kumar. 1926. *The origin and development of the Bengali language*. Calcutta: Mehra, Rupa.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Clark, Robin & Ian Roberts. 1993. A computational model of language learnability and language change. *Linguistic inquiry* 24(2):299–345.
- Comrie, Bernard. 1978. Ergativity. In W.P. Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*, 329–394. Austin: University of Texas Press.
- Dahl, Eystein & Krzysztof Stroński. 2016. Ergativity in Indo-Aryan and beyond. In *Indo-Aryan ergativity in typological and diachronic perspective*, 1–38. Amsterdam: John Benjamins.
- Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.

- DeLancey, Scott. 1981. An interpretation of split ergativity and related patterns. *Language* 626–657.
- Deo, Ashwini & Devyani Sharma. 2006. Typological variation in the ergative morphology of Indo-Aryan languages. *Linguistic Typology* 10:369–418.
- Dixon, R. M. W. 1979. Ergativity. *Language* 55:59–138.
- Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511611896.
- Dresher, Bezalel Elan. 1999. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30(1):27–67.
- Galbraith, Daniel. 2022. *Optimal linking grammar: A theory of morphosyntax*. Cambridge: Cambridge University Press.
- Gavrilets, Sergey. 2004. *Fitness landscapes and the origin of species*. Princeton: Princeton University Press.
- Gelderen, Elly van. 2023. *The linguistic cycle: Economy and renewal in historical linguistics*. London: Routledge.
- Genetti, Carol. 1993. On the morphological status of casemarkers in Dolakha Newari. *Linguistics of the Tibeto-Burman area* 16(1):57–73.
- Gündoğdu, Songül. 2017. Variation in the Ergative pattern of Kurmanji. *Wiener Jahrbuch für Kurdische Studien* 5:45–62.
- Gündoğdu, Songül. 2022. Case in Kurdish. In Yaron Matras, Geoffrey Haig & Ergin Öpengin (eds.), *Structural and typological variation in the dialects of Kurdish*, 111–180. Cham: Palgrave Macmillan.
- Gusain, L. 2000. *Bagri*. München: Lincom Europa.
- Haider, Hubert. 2023. SVO attractor in the declarative-to-procedural shift in grammar evolution. *Acta Linguistica Academica* 70(2):195–218. doi:10.1556/2062.2023.00642.
- Haig, Geoffrey. 2008. *Alignment change in Iranian languages. a Construction Grammar approach*. de Gruyter: Oxford University Press.
- Haig, Geoffrey. 2017. Deconstructing Iranian ergativity. In Jessica Coon, Diane Massam & Lisa Demena Travis (eds.), *The Oxford handbook of ergativity*, 465–500. Oxford: Oxford University Press.
- Halle, Morris. 1962. Phonology in generative grammar. *Word* 18:54–72.
- Hawkins, J. A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

- Jakobson, Roman. 1929. Remarques sur l'évolution phonologique du Russe comparée à celle des autres langues slaves. *Travaux du Cercle Linguistique de Prague* 2. Reprinted in *Selected Writings*, Vol. 1.
- Jakobson, Roman. 1941. *Kindersprache, Aphasie, und allgemeine Lautgesetze*. Uppsala: Uppsala Universitets Årsskrift.
- Jamison, Stephanie. 2000. Lurching towards ergativity: Expressions of agency in the Niya documents. *Bulletin of the School of Oriental and African Studies* 63(1):64–80. <http://www.jstor.org/stable/1559588>.
- Jasbi, Masoud. 2020. The meaning of the Persian object marker *rā*: What it is not, and what it (probably) is. In *Advances in Iranian linguistics*, 119–136. Amsterdam: Benjamins.
- Joshi, Kinjal. 2020. Agreement and information structure in Surati Gujarati. In Christopher Pinon & Laurent Roussarie (eds.), *Empirical issues in syntax and semantics 13 (eiss 13)*, 1–24. Paris: CSSP.
- Jügel, Thomas. 2015. *Die Entwicklung der Ergativekonstruktion im Alt- und Mittelindischen*. Wiesbaden: Harrassowitz.
- Kim, Ronald I. 2025. Case marking from Old Iranian to Ossetic: A comparative and typological approach. *Transactions of the Philological Society* 123(2):280–313. doi:10.1111/1467-968X.12316.
- Kiparsky, Paul. 1993. An OT approach to phonological variation. Rutgers Optimality Workshop (also presented at NWave 1993). <https://web.stanford.edu/~kiparsky/Papers/nwave94.pdf>.
- Kiparsky, Paul. 2002. Event structure and the perfect. In David I. Beaver, Luis D. Casillas Martínez, Brady Z. Clark & Stefan Kaufmann (eds.), *The construction of meaning*, CSLI Publications.
- Kiparsky, Paul. 2012a. Grammaticalization as optimization. In Dianne Jonas, John Whitman & Andrew Garrett (eds.), *Grammatical change: Origins, nature, outcomes*, Oxford: Oxford University Press.
- Kiparsky, Paul. 2012b. Greek anaphora in cross-linguistic perspective. *Journal of Greek Linguistics* 12:84–117.
- Kiparsky, Paul. 2026. Syntactic change: The word-order cycle. In Jadranka Gvozdanović (ed.), *Proceedings of ICHL 25*, Amsterdam: Benjamins.
- Lahiri, Bornini. 2021. *The case system of eastern Indo-Aryan languages: A typological overview*. Oxford and New York: Routledge.
- Li, Chao. 2007. Split ergativity and split intransitivity in Nepali. *Lingua* 117:1462–1482.
- Lindemann, Luke. 2024. The syntactic profile of Nepali ergativity. *Nepalese Linguistics* 39(1):49–59. DOI:<https://doi.org/10.3126/nl.v38i1.71557>.

- Liperovskij, V.P.. 1988. *Očerki grammatiki sovremennogo bradža [an outline of the grammar of contemporary Braj]*. Moskva: Izdatel'stvo Nauka.
- Magier, David. 1983. *Topics in the grammar of Marwari*: Berkeley dissertation.
- Masica, Colin P. 1993. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- McConvell, Patrick. 1981. How Lardil became accusative. *Lingua* 55(2):141–179. doi: [https://doi.org/10.1016/0024-3841\(81\)90061-9](https://doi.org/10.1016/0024-3841(81)90061-9). <https://www.sciencedirect.com/science/article/pii/0024384181900619>.
- Meillet, Antoine. 1918. Convergence des développements linguistiques. *Revue philosophique* 85:97–110.
- Mistry, P. J. 1997. Objecthood and specificity in Gujarati. In Jane Hill, P. J. Mistry & Lyle Campbell (eds.), *The life of language*, 425–442. Berlin: Mouton de Gruyter.
- Mistry, P.J. 2004. Subjecthood of non-nominatives in Gujarati. In Peri Bhaskararao & Karumuri V. Subbarao (eds.), *Non-nominative subjects*, vol. 2. Typological Studies in Language, 1–32. Amsterdam and Philadelphia: Benjamins.
- Mohanan, Tara. 1989. Syllable structure in Malayalam. *Linguistic Inquiry* 20(4):589–625.
- Mohanan, Tara. 1994. *Argument structure in Hindi*. Stanford, CA: CSLI Publications.
- Montaut, Annie. 2018. The rise of differential object marking in hindi and related languages. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking* (Studies in Diversity Linguistics 19), 253–282.
- Mosegaard Hansen, Maj-Britt & Richard Waltereit. 2025. Cyclic change in grammar and discourse.
- Patel-Grosz, Pritty. 2021. Ergativity in Indo-Aryan. In *Oxford research encyclopedia of linguistics*, Oxford: Oxford University Press.
- Riggle, Jason. 2009. The complexity of ranking hypotheses in Optimality Theory. *Computational Linguistics* 35(1):47–59.
- Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt Brace & World.
- Seržant, Ilja A. & George Moroz. 2022. Universal attractors in language evolution provide evidence for the kinds of efficiency pressures involved. *Humanities and Social Sciences Communications* 9:58. <https://doi.org/10.1057/s41599-022-01072-0>.
- Stroński, Krzysztof. 2009. Approaches to ergativity in Indo-Aryan. *Lingua Posnaniensis* 51(1):77–118.
- Stroński, Krzysztof. 2010. Variation of ergativity patterns in Indo-Aryan. *Poznań Studies in Contemporary Linguistics* 46(2):237–253. doi:10.2478/v10010-010-0012-8.

- Verbeke, Saartje. 2013. *Alignment and ergativity in new Indo-Aryan languages*, vol. 51. Berlin/Boston: Mouton de Gruyter.
- Verbeke, Saartje & Eva De Clercq. 2016. Looking for ergativity in Indo-Aryan. In Eystein Dahl & Krysztof Stroński (eds.), *Indo-Aryan ergativity in typological and diachronic perspective*, 39–61. Amsterdam: John Benjamins.
- Wexler, Kenneth & Rita Manzini. 1987. Parameters and learnability in binding theory. In *Parameter setting*, 41–76. Springer.
- Witzlack-Makarevich, Alena & Ilja A. Seržant. 2018. Diachrony of differential argument marking. In Alena Witzlack-Makarevich & Ilja A. Seržant (eds.), *Differential argument marking: Patterns of variation* (Studies in Diversity Linguistics 19), chap. 1:1–40. Berlin: Language Science Press.
- Yang, Charles. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.
- Yang, Charles D. 2000. Internal and external forces in language change. *Language variation and change* 12(3):231–250.