Modeling syntactic drift: Indo-Iranian case and agreement

Major linguistic changes tend to occur in recurrent types of trajectories by small steps over extended periods, a phenomenon known as DRIFT. The challenge for historical and general linguistics is to explain the unidirectionality of drift, its discrete incremental (rather than gradual or "catastrophic") actuation, its grounding in grammatical theory, and its relation to typology and synchronic variation. We present a model of syntactic drift, and show that, coupled with a theory of structural case, it accounts for the stepwise dismantling of the Middle Indic ergative case-marking system in the modern Indic languages (Deo & Sharma 2006), and for the corresponding drift in Iranian, which runs independently on a largely parallel track (Haig 2008). The main changes are these: (1) A new accusative case is extended to definite/human objects in all Indo-Aryan languages (originally a dative), and in some Iranian languages (grammaticalized from various sources). (2) When the subject is ergative, the new accusative objects continue to agree with the verb in some languages (Gujarati); in others, the verb then gets 3P Masc. default agreement (e.g. Hindi, Marathi). (3) Agreement is transferred to ergative subjects, regardless of whether the object is accusative or nominative (e.g. Nepali, Eastern Pahāḍī, Northern Kurdish). (4) Ergative case is lost (e.g. Bengali, Oriya, Bhojpuri, Farsi, Gilan Kurdish). The attested grammars represent different stages in this trajectory, and different orders of the individual changes that constitute it.

We start from the assumption that a language's course of historical development is shaped by systematic preferences that govern learners' choices among alternative grammars consistent with the data encountered. Specifically, we suppose that a learner at any given stage of acquisition posits the most probable language consistent with the data. The theory is consistent with several specific models of grammar and acquisition, but we implement it in because OT provides a simple formal measure of the probability of a language: its RANKING VOLUME, defined as the number of constraint rankings that generate it, divided by $k!$, where k is the number of constraints (Riggle 2010). Ranking volume is already known to be a a good predictor of typological distribution (Bane & Riggle 2008) and of relative frequency of free variants (Kiparsky 1993, Anttila 1997, 2010). We use PyPhon to compute the typology and the ranking volumes.

We proceed by constructing OT analyses for the subsystem of interest in a family of languages. We posit that the constraints are universal, and, initially, that all rankings are possible and equiprobable. We pool the constraints that are active in each of the languages, generate the factorial typology, and compute the ranking volumes of the languages in the typology. The prediction is that these characterize the the possible historical pathways within the typological space.

The following illustrative tableau shows the generation of a perfective transitive clause with definite subject and object in Hindi/Urdu. In the input, subject and object are respectively specified for abstract Ergative [–LR,+HR] and Accusative [+LR,–HR], by universal rules on the basis of argument structure. Undominated constraints (omitted from the tableau, along with the candidates that violate them) require that only nominals get case, that only verbs agree, and that agreement is with exactly one nominal (possibly expletive). The rankings $\text{MAX}_{\text{Erg}} > *\text{ERG}$ and $\text{MAX}_{\text{Acc}} > *\text{ACC}$ ensure that the abstract Ergative and Accusative cases of the subject and object are mapped into the corresponding morphosyntactic cases. The high ranking of $*\text{AGR}_{\text{ERG}}$ and $*\text{AGR}_{\text{Acc}}$ prevents the verb from agreeing with either of them, so that a null nominative expletive subject is inserted for the verb to agree with. Currently the system models aspectually conditioned split ergative case marking, and definiteness/animacy-sensitive differential object marking. (It is easily extensible to the splits between nominal and pronominal case marking found in some of the languages.)

	Hindi/Urdu	A *AGR$_{ERG}$	B *AGR$_{Acc}$	C MAX$_{Erg}$	D MAX$_{Acc}$	E *ACC	F *ERG	G AGR$_{Nom}$	H AGR$_{Hi}$	I *∅
	/Erg$_{Def}$Acc$_{Def}$V$_{Perf}$/									
1.	Erg Acc Agr$_{Erg}$	*				*	*			
2.	Erg Acc Agr$_{Acc}$		*			*	*		*	
☞ 3.	∅ Erg Acc Agr$_{∅}$					*	*			*
4.	Erg Nom Agr$_{Erg}$	*			*		*	*		
5.	Erg Nom Agr$_{Nom}$				*		*	*	*	
6.	∅ Erg Nom Agr$_{∅}$				*		*			*
7.	Nom Acc Agr$_{Nom}$			*		*				
8.	Nom Acc Agr$_{Acc}$		*	*		*		*	*	
9.	∅ Nom Acc Agr$_{∅}$			*		*		*		*
10.	Nom$_x$ Nom$_y$ Agr$_x$			*	*					
11.	Nom$_x$ Nom$_y$ Agr$_y$			*	*				*	

These nine constraints yield 9! = 362,880 rankings, which converge on just ten distinct languages. All but two of them, with very low *r*-volumes (<0.01), correspond to attested languages. Bangla, Farsi (Nom/Acc) and Central Kurdish (no case) have the highest *r*-volumes (0.29). These are the end points of the development, derivable via a sequence of minimal rerankings that yield different intermediate varieties of ergative case systems. The analysis thus accounts both for the attested direction of drift, and for the alternative attested pathways through a multidimimensional fitness landscape: spread of accusative object, spread of nominative subject, loss of object agreement and spread of subject agreement.

Importantly, the formal markedness and faithfulness constraints on accusative and ergative case are entirely parallel. The drift is generated by the two constraints that require agreement with the nominative argument and with the most prominent argument (AGR$_{Nom}$ and AGR$_{Hi}$ in the above tableau). In fact, the same constraints within a different case system account for syntactic variation and change in Insular Scandinavian. The shift from the Icelandic construction of dative-subject verbs with nominative objects that trigger number agreement to Faroese accusative objects and default third person singular agreement is analogous to the shift from Middle Indic to Hindi/Marathi, and involves the same constraint reranking (Galbraith 2018).

Consequently, under this set of constraints, the theory predicts that ergative case systems should be more stable in languages that have no subject-verb agreement than in languages that do. The model also accounts for the important implicational generalization that replacement of ergative subjects by nominative subjects always presupposes transfer of agreement from nominative objects to subjects: agreement is never "more ergative" than case marking (Haig 2008: 304). Grammars that violate this generalization are harmonically bounded, that is, underivable on any constraint ranking.

The diachronic picture that emerges is that long-term change is a transition between canonical systems, empirically identifiable by their typological frequency and relative historical stability, and formally characterizable as *r*-volume maxima. Non-canonical (marked) systems necessarily arise as transitional stages between canonical systems.

References

ANTTILA, ARTO. 1997. Deriving variation from grammar. In Frans Hinskens, Roeland van Hout, & Leo Wetzels (eds.), *Variation, Change and Phonological Theory*. Amsterdam / Philadelphia: John Benjamins, pp. 35-68.

ANTTILA, ARTO. 2015. Free Variation in Finnish Structural Case. https://web.stanford.edu/~anttila/research/lsa-2015-handout.pdf

BANE, MAX & JASON RIGGLE . 2010. The typological consequences of weighted constraints. *CLS* 45.

BOWMAN, SAM. Building OT Grammars in PyPhon. 2012 https://www.nyu.edu/projects/bowman/pinterest-pyphon-tutori pdf

DEO, ASHWINI & DEVYANI SHARMA. 2006, Typological variation in the ergative morphology of Indo-Aryan languages. *Linguistic Typology* 10:3.

GALBRAITH, DANIEL. 2018. *The Predictable Case of Faroese.* PhD Dissertation, Stanford University.

HAIG, GEOFFREY. 2008. *Alignment Change in Iranian Languages: A Construction Grammar Approach.*

KIPARSKY, PAUL. 1993. Variable rules. Rutgers Optimality Workshop 1993, and NWAVE 1994. http://www.stanford.edu/~kiparsky/Papers/nwave94.pdf