Some universals of reflexive construction markers
and a possible efficiency-based explanation

Martin Haspelmath
Max Planck Institute for the Science of Human History (Jena) / Universität Leipzig

Many of the world’s languages have special reflexive construction markers, which come in two main types (Faltz 1977): verbal REFLEXIVE VOICE MARKERS, and REFLEXIVE NOMINALS (often called “reflexive pronouns” or “anaphors”). For example, Finnish has the suffix -utu- (e.g. riisu-utu-a ‘undress (oneself)’), and Persian has the reflexive nominal xod (e.g. u xod-rā košt [he self-ACC killed] ‘he killed himself’). While many linguists have investigated the syntactic conditions for the occurrence of reflexive nominals (vs. nonreflexive pronouns) (e.g. Gair et al. 2000, among many others), and while the syncretism patterns of reflexive voice markers have often been studied (e.g. Kemmer 1993), there has not been much cross-linguistic research comparing the two construction types. And quite generally, while much work has focused on language-particular analyses, the forms of the reflexive construction markers have rarely been studied in a broad cross-linguistic perspective (but see Schladt 1999; König et al. 2005 for reflexive nominals derived from body-part terms and self-intensifiers).

Against this background, the present paper reports on a study of 50 languages worldwide, from 50 unrelated language families, in which I found evidence supporting the following universal hypotheses:

1. If a language has a reflexive voice marker, one of its uses is for agent-patient coreference.
2. If a language uses different constructions for agent-patient coreference for different verb types, then it uses shorter markers for introverted verbs than for extroverted verbs (cf. König & Vezzosi 2004).
3. In all languages, the usual coding of disjoint anaphoric reference is at least as short as the usual coding of agent-patient coreference (cf. Haspelmath 2008: 48).
4. If a language has nonreflexive bound person forms for objects, these cannot be used subject-coreferentially in agent-patient contexts.
5. If a language uses different reflexive construction markers for object function and adnominal possessor function, then the adnominal possessor marker is shorter than the object marker.

The research strategy of looking for universal generalizations rather than focusing on language-particular analyses is motivated by the recognition that comparative studies need not be based on in-depth analyses of particular languages, but can be based on surveying comprehensive grammatical descriptions (as in Dryer 2018 and much other work in this vein). Universals found in this way may be explainable in a variety of ways (cf. Schmidtke-Bode at el. 2019), including innate categories, diachronic tendencies, and efficiency principles (Gibson et al. 2019). While not excluding an explanation in terms of innate categories, in this paper I will explore the hypothesis that universals (2)-(5) can be explained
by a general Zipfian principle of efficient coding: Greater predictability results in shorter forms. I will try to show that the shorter forms also tend to occur in contexts where their reference is more predictable. Finally, I will briefly argue against explanations that rely exclusively on diachronic tendencies. I currently lack an explanation of the first universal, but I include it here because my primary goal is to find universals, regardless of what might explain them.

Appendix

The 50 languages of my sample are: Balanta-Ganja, Bardi, Basque, Bininj Gun-wok, Burushaski, Cavineña, Chiapas Zoque, Coptic, English, German Sign Language, Haida, Hup, Itzaj, Kalaallisut, Kamaiurá, Kayardild, Komnzo, Korean, Koyra Chiini, Krongo, Laguna Keres, Lavukaleve, Lezgian, Ma’di, Mandarin Chinese, Mandinka, Mapudungun, Maricopa, Martuthunira, Mauwake, Maybrat, Motuna, Murui, Nenets, Nzadi, Oklahoma Cherokee, Panare, Rapa Nui, Sandawe, Savosavo, Sheko, Teiwa, Ts’ixa, Ulwa, Ute, Wambaya, Wappo, Yauyos Quechua, Yukaghir, Yurakaré

References