

The other end of universals: theory and typology of *rara*

Michael Cysouw & Jan Wohlgemuth

1 *Rara* and *Rarissima*

Universals of language have been studied extensively for at least the last four decades, allowing fundamental insights into the principles and general properties of human language. Only incidentally have researchers looked at the other end of the scale. And even when they did, they mostly just noted peculiar facts as “quirks” or “unusual behavior”, without making too much of an effort at explaining them beyond calling them exceptions to various rules or generalizations. Systematic accounts of *rara* are scarce, however, see the brief overview on page 5.

Yet, *rara* and *rarissima*, features and properties found in very few languages, can tell us as much about the capacities and limits of human language(s) as do universals. Explaining the existence of cross-linguistically rare phenomena on the one hand, and the fact of their rareness or uniqueness on the other, should prove a reasonable and interesting challenge to any theory of how human language works. The current volume consists of papers dealing with such rarities, their analysis, and their impact on the study of human language in general.

A *rarum* (and its extreme case, a *rarissimum*) is not just something that is rare or infrequently attested. In the introduction to his “Raritätenkabinett”,¹ Plank defines a *rarum* as

“... a trait ... which is so uncommon across languages as not even to occur in all members of a single ... family or diffusion area ... Diachronically speaking, a *rarum* is a trait which has only been retained, or only been innovated, in a few members of a single family or sprachbund or of a few of them.”

With this definition, Plank very specifically delimits a *rarum* from other infrequent phenomena among the world’s languages. Following Plank, a *rarum* should not just be infrequent, but its attestations should also be independent, i. e. it should also never occur locally spread out, forming either genealogical and / or geographical clusters.

A similar view of *rara* is formulated by Bickel and Nichols (2003: 3). They distinguish between two types of *rara* that are rather different in their quality. The first type, *absolute rara*, are those that are found rarely across language families and thus *rara* in Plank's sense. One example of this type of *rara* is found in the languages Pirahã and Kawi which have no number distinction in pronouns, thus effectively violating the Greenbergian universal 43 (cf. Frerick 2006: 41; Greenberg 1963: 113). The second type, *relative rara*, are those that are rare on a global scale but common within a geographical area or a language family. A prime example for this type are click phonemes: Their distribution is restricted to Southern and Eastern Africa, where they are common among several, yet not all, groups of languages, while clicks are essentially unattested in all other parts of the world — and thus relatively rare on a global scale (cf. Frerick 2006: 10, 68).²

Plank (2000) suggests a few other terms for talking about rare phenomena. He proposes the term *singulare* for features found in only one language, but this term has an inherent problem when used in English: the adjective derived from it is homophonous with the noun and adjective referring to grammatical number category SINGULAR (as opposed to e. g. PLURAL). In a similar vein, *nonetheless*, the alternative term for *singulare* suggested by Plank (2000), might evoke the false interpretation that there were *no* language with such a characteristic. Furthermore, this term bears the connotation of a value judgment since *nonetheless* also means 'someone or something that is better than all others'. To avoid homonymous or misleading terms, we prefer not to adopt these terms but suggest to use *unicale / unique* instead for such features that apparently are attested in only one language. Whatever term one prefers, it is of course to a large extent only of superficial interest that there is just and exactly *one* single known example of a particular phenomenon. The study yielding this one example will only have looked at a limited set of other languages — enlarging the sample of languages might very well turn up more cases. Thus, absolute numbers of occurrence never tell very much about the prevalence of a characteristic among the world's languages.

For the sake of brevity some linguists use the collocations "*rare language(s)*" and "*unique language(s)*" to refer to languages *having* such rare or unique characteristics. This, however, seems inappropriate to us, especially in the context of language endangerment,³ and given the fact that, by virtue of its specific combination of features and characteristics, *every* language is unique.

2 The study of *rara*

A central goal of investigating *rara* is to fathom the variability and limits of human language structure(s). Broad-scale typological research using samples of the world's languages will give an indication about what are the common kinds of linguistic structures. Yet, such studies will not be able to accurately depict the fringes of human languages, i. e. those structures that are only rarely attested. Far too often, these rare structures are hidden in a heterogeneous waste-basket category of unclassifiable 'other' structures in typological surveys.

Admittedly, the search for, and study of, *rara* is methodologically difficult. There is no principled method for studying objects that are only rarely attested, except for using extremely large samples (which is normally too labor-intensive to be practically feasible). The only option seems to be to rely on serendipitously noted cases — either as a by-product of large-scale typological surveys or stemming from specific descriptions of mystifying phenomena encountered by specialists of a particular language. Starting from such a nucleus of known cases, the search for similar phenomena can be continued through checking closely related languages and areally close languages. Still, such a search for *rara* inevitably takes time, and the research will often span many years (or even decades) as a side-track of other research activities.

On the basis of the current knowledge about the diversity of human languages it remains infeasible to decide whether unattested structures are absolutely impossible or simply highly improbable. We presently “only” have some knowledge about a few thousand languages, and the variability of these languages is highly constrained by genealogical and areal cohesion. The fact that something is not attested among the sufficiently described world's languages might thus just as well be the result of historical coincidences instead of a sign of limits on the structural possibilities of human language.

Explicitly studying rarities will present a much more detailed picture of what is linguistically possible. An excellent example of the importance of studying *rara* for the understanding of the limits of the structure of human language is the paper on the interaction between gender and number by Plank and Schellinger (1997). They start from the well known Greenbergian (1963) typological universals 37 and 45, which state that gender distinctions in the plural imply gender distinctions in the singular. However, Plank and Schellinger show that – on closer inspection – a large set of “counterexamples” exists. Instead of considering such counterexamples nuisance elements that

spoil an otherwise nice theory or generalization, Plank and Schellinger argue that these counterexamples be taken as opportunities: by collecting and interpreting such “exceptional” examples, a deeper and more accurate understanding of the possible variability of human language can be reached.

A different goal of the study of *rara* and *rarissima* is to argue against widespread assumptions about the limits of possibilities of human language. Either some generalizations had been proposed to which “counterexamples” are attested (like in the case of the correlation between genders and numbers discussed above), or some phenomenon that was deemed to be completely impossible is shown to exist after all. A prominent example of this kind of study is the survey of the labial flap by Olson and Hajek (2003). This sound, the only non-rhotic flap, has long been thought to be non-existent or at least not to be a distinctive phonological unit in any language. Yet, as Olson and Hajek (2003) showed, the labial flap exists in about 70 languages of Africa and one in Indonesia and in 22 of these languages the sound is indeed a distinctive unit contrasting with other bilabials.

Yet another possible use of *rara* is in tracking historical connections between languages. If any set of languages shares a rare or unique feature or even a bundle of “shared quirks”, this is a strong indicator for a shared history of ancient contact or common descent, making these occurrences a useful diagnostic in diachronic linguistics and typology. This has e. g. been illustrated by Gensler (1994, 1997, 2003) by using different syntactic parameters and constructions as evidence for ancient language contact. For example, the syntagm S-AUX-O-V-OTHER can be reconstructed for Proto-Niger-Congo and is common all over the family. The same sequence is, however, basically unattested outside the family apart from half a dozen languages of Sudan into which it must have diffused.

In general though, the main question raised by the existence of *rara* is how to deal with them in theoretical approaches to language. The fact that *rara* exist – and even stronger, that the existence of *rara* as such does not seem to be exceptional at all – suggest that a theory of linguistic structure should have some principled notion of dealing with the existence of rare traits of human languages. Cysouw (2005: 248) estimates for person-marking syncretisms that even when taking the somewhat more widespread *rara* into account in a theory, there still are about 16% of the world’s languages that possess some structure which is rare. Each of these cases in itself is a *rarum*, but all together they make up a sizable portion of the world’s linguistic structures. So, it does not suffice to simply dismiss any *rara* as incidental aberrations in the space-

time of linguistic structure, as “exceptions” or “historical coincidences”. The real challenge is to develop theoretical notions for human language that inherently can deal with rarity and other types of variation.

At any rate, the terms *rarum* and *rarissimum* are used to refer to grammatical characteristics found only in very few languages, where the latter term would be referring to characteristics found in even fewer of the world’s languages. For a more tangible quantification, a threshold of attestations in $\leq 5\%$ of the world’s languages for *rara* and in $\leq 1\%$ of the world’s languages for *rarissima* has been discussed by Frerick (2006: 65–67), noting that such quantification is rather arbitrary. One must bear in mind that $\leq 1\%$ of about 7,000 languages still amounts to approximately 70 languages on a worldwide basis. And, given that the current world’s languages can be grouped into about 350 different genera (Dryer 2005), the criterium of non-genealogical clustering of *rara* would result in each fifth genus having a language with the *rarissimum* in question. From this perspective, even the $\leq 1\%$ criterium does not seem that unusual after all. Yet, with not even half of these about 7,000 languages being properly described, an attestation of a characteristic in $\leq 1\%$ of the world’s *documented* languages already boils this figure down to probably less than 30 languages worldwide.

A different take on defining *rara* is to try and establish the stability of a linguistic phenomenon through time. The underlying rationale of Plank’s definition of *rara* (viz. *absolute rara* in the Bickel and Nichols sense) is that a *rarum* is a phenomenon that could very well arise in a particular language (after all, languages allow all kinds of strange things to happen), but when this happens it should not be for too long. The *rarum* should be an ‘instable’ characteristic and quickly change again into something else. Reformulating this idea as a dynamic process, it suggests that the possibility of ‘change away’ from a *rarum* to something else should be much greater than the probability of the *rarum* arising in the first place. As a measure of rarity one could then use the quotient of the probability of a ‘change away’ and the probability of the *rarum* arising. As a matter of fact, though, at least some *relative rara* appear to be extremely stable and can even be traced back to ancestral languages, as noted e. g. by Harris (this volume: 98). This question suggests that the study of *rara* should be of great interest to the investigation of the dynamics of language change and vice versa.

Compared to the ongoing research tradition on language universals it is rather recently that there are various activities dealing with rarities among the world’s languages. First and foremost there is “*das grammatische Raritäten-*

kabinett: a leisurely collection to entertain and instruct” already mentioned above which has been edited and published online for more than a decade now by Frans Plank. This easily searchable database is a good starting point for any investigation into rare or infrequent structures of human languages.

Furthermore, in the same time frame in which the *Rara & Rarissima* conference and this volume were prepared, Horst Simon and Heike Wiese organized a session during the 27th annual meeting of the Deutsche Gesellschaft für Sprachwissenschaft in Cologne (DGfS Jahrestagung 2005), entitled “Expecting the Unexpected — Exceptions in Grammar”. This session will also result in a collection of papers (Simon and Wiese (eds.) forthcoming). Although the topic of exceptions is not necessarily the same as the study of rarities, there is still a good chance that rarities will be unexpected and occasionally even overlooked exceptions with respect to many theoretical proposals about the structure of human language.

3 Survey of this book

This book consists of various papers dealing with the theory and / or typology of *rara* among the world’s languages. There is also a companion volume to the present book dealing with the details of rare and unusual structures in individual languages, namely “*Rara & Rarissima: Documenting the fringes of linguistic diversity*” (Wohlgemuth and Cysouw (eds.) 2010).

The current volume starts with two papers dealing with numeral systems among the world’s languages, the first by *Harald Hammarström* “Rarities in numeral systems” and the second by *Thomas Hanke* “Additional rarities in the typology of numerals”. Numeral systems have a long history of typological investigations (see the references in these papers), so this domain of linguistic structure is a prime example in which the study of *rara* can supplement known general tendencies with lesser-known minor tendencies.

The paper by *Alice Harris* “Explaining typologically unusual structures: The role of probability” is the first of various papers in this volume dealing explicitly with the challenge that *rara* pose for theoretical consideration of language structure (see also the papers by Malchukov, Newmeyer, and Rijkhoff). Harris argues that *rara* are rare because it is unlikely for them to arise. Specifically, she illustrates this by rare phenomena that only arise through a combination of various diachronic steps. Each change individually is not necessarily special in any sense, but the combination of all diachronic requirements makes the end result unusual from a world-wide perspective.

Taking Plank's definition of *rara* seriously, the paper by *Pavel Iosad* "Right at the left edge: initial consonant mutations in the languages of the world" is not really about a *rarum*. As he shows, initial consonant mutation is incidentally found throughout the world's languages, but it is also a general trait of the Celtic languages. Such a consistent distribution throughout all members of a genealogical group shows that although the trait might be unusual from a worldwide perspective, it is still a stable possibility for a language to portray and does not count as a real *rarum*. The paper by Iosad can thus be read as (implicitly) arguing that initial consonant mutation is not a *rarum* in Plank's sense after all, but rather a relative *rarum* in Bickel and Nichols' sense.

Various possible explanations for rarities and rareness are presented by *Andrej Malchukov* in his paper "Quirky case: Rare phenomena in case-marking and their implications for a theory of typological distributions". Malchukov describes a few unusual phenomena related to case marking. These examples illustrate three different reasons why a phenomenon might be a *rarum*. First, a rare pattern may result from a conflict between a grammaticalization path and a functional constraint. Second, a pattern may be rare as it requires the co-occurrence of several different conditions. And third, functionally deviant cases may result from incomplete grammaticalization cycles.

In his paper "Negatives without negators" *Matti Miestamo* takes up the challenge of a long-known typological (relative) *rarum*: the marking of negation by the absence of linguistic marking in some Dravidian languages. He compares the situation in such languages to the world-wide diversity of the marking of negation, pointing out various partial parallels in other languages. By combining the typological survey with the study of a *rarum*, Miestamo is able to make some sense of the otherwise rather puzzling negation structure in Dravidian.

The next two papers take the central question of *rara* head-on: how should *rara* be treated by theoretical notions of language structure. *Frederick J. Newmeyer* notes in his paper "Accounting for rare typological features in formal syntax: Three strategies and some general remarks" that rarities present a particular challenge for the Principles & Parameters approach to language, given the central idea of this approach that seeming complexity and idiosyncrasy are purely epiphenomenal. He argues that the existence of a rare feature is derivable from the interaction of processes known to be motivated in the grammars of the world's languages.

Jan Rijkhoff in his paper “Rara and grammatical theory” discusses various *rara* in the domain of noun phrase structure in the context of *Functional Discourse Grammar*. More generally, though, he argues that *rara* play a crucial role in the validation of claims made by any theory.

The question how to quantify the overall level of rarity of a language is taken up by *Søren Wichmann and Eric W. Holman* in their paper “Pairwise comparisons of typological profiles”. Using the *World Atlas of Language Structures* and computing degrees of (typological) difference between two languages at a time, they investigate the relation between genealogical relationship and typological profiles of languages.

Finally, the paper by *Jan Wohlgemuth* “Some reflections on the interrelation of language endangerment, community size and typological rarity” investigates the influence of non-linguistic characteristics of a speaker community on *rara*. Specifically, he argues that there is a relation between the overall rarity of a language and its endangerment status.

Notes

1. <http://typo.uni-konstanz.de/rara/intro/index.php?pt=1>
2. Clicks were, however, also attested independently in the extinct speech register Damin of Lardil in Australia (cf. Hale 1998: 204 *passim*)
3. cf. Wohlgemuth (this volume)

References

- Bickel, Balthasar and Johanna Nichols
2003 *Typological enclaves*. Paper presented at the 5th Biannual Conference of the Association for Linguistic Typology (ALT V), Cagliari, September 18.
<http://www.uni-leipzig.de/~autotyp/download/enclaves@ALT5-2003BB-JN.pdf>
- Cysouw, Michael
2005 What it means to be rare: the case of person marking. In *Linguistic Diversity and Language Theories*. Zygmunt Frajzynger, Adam Hodges and David S. Rood (eds.), 235–258. Amsterdam: Benjamins.
- Cysouw, Michael
forthc. *Quantitative explorations of the world-wide distribution of rare characteristics, or: the exceptionality of north-western European languages*. To appear in *Exception in Language*, Horst Simon and Heike Wiese (eds.). Berlin / New York: Mouton de Gruyter.

- Dryer, Matthew
2005 *Genealogical Language List*. In: Martin Haspelmath, David Gil, Matthew S. Dryer and Bernard Comrie (eds.): *The World Atlas of Language Structures*. Oxford etc.: Oxford University Press, 584–643.
- Frerick, Daniela
2006 *Raritäten in den Sprachen der Welt*. Westfälische Wilhelms-Universität Münster, M.A. thesis.
- Gensler, Orin
1994 On reconstructing the syntagm S-Aux-O-V-Other to Proto-Niger-Congo. In *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistic Society, February 18-21, Special Session on Historical Issues in African Linguistics (BLS 20-S)*, Kevin E. Moore, David A. Peterson, Comfort Wentum (eds.), 1–20. University of California, Berkeley.
- Gensler, Orin
1997 Grammaticalization, typology, and Niger-Congo word order: Progress on a still-unsolved problem. Review article on “Die Stellung von Verb und Objekt in Niger-Kongo-Sprachen”, by Ulrike Claudi. *Journal of African Languages and Linguistics* 18 (1): 57–93.
- Gensler, Orin
2003 Shared quirks: A methodology for “non-orthodox” historical linguistics. Paper presented at the 17th International Congress of Linguists, Prague, 29 July 2003.
- Greenberg, Joseph H.
1963 Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, Joseph H. Greenberg (eds.), 73–113. Cambridge, Mass.: MIT Press.
- Hale, Ken
1998 On endangered languages and the importance of linguistic diversity. In *Endangered Languages. Current issues and future prospects*, Lenore A. Grenoble and Lindsay J. Whaley (eds.), 192–216. Cambridge etc.: Cambridge University Press.
- Olson, Kenneth S. and John Hajek
2003 Crosslinguistic insights on the labial flap. *Linguistic Typology* 7 (2): 157–186.
- Plank, Frans and Wolfgang Schellinger
1997 The uneven distribution of genders over numbers: Greenberg Nos. 37 and 45. *Linguistic Typology* 1 (1): 53–101.
- Simon, Horst and Heike Wiese (eds.)
forthc. *Exception in Language*. Berlin / New York: Mouton de Gruyter.
- Wohlgemuth, Jan and Michael Cysouw
2010 *Rara & Rarissima: Documenting the fringes of linguistic diversity*. Berlin / New York: Mouton de Gruyter. (= Empirical Approaches to Language Typology; ##). ¹