[Revised version April 2025, to appear in a Festschrift]

PHENOMENON-BASED LANGUAGE SAMPLING

Matti Miestamo

Some linguistic phenomena, such as egophoricity or vowel harmony, have a clearly restricted areal or genealogical distribution in the languages of the world. In typological studies of such phenomena, the sample languages are often selected among languages that are known, on the basis of earlier research, to exhibit the phenomenon under study, rather than using an extensive sample with a balanced world-wide representation of languages. The paper proposes to call this approach to language sampling *phenomenon-based sampling*. After a short overview of types of language sampling in typology, the notion of phenomenon-based language sampling is introduced and defined. A survey of relevant articles in the journal *Linguistic Typology* 1997–2024 is made to form a general picture of the kinds of phenomenon-based solutions to sampling that have been used in typological studies over the last thirty years. Attention is paid to the ways of identifying the languages that exhibit the phenomenon and to the ways of selecting the sample languages from among the languages thus identified. The various methodological solutions are compared, and interesting similarities and differences are taken up for discussion. It is recommended that instead of downplaying the role of sampling by characterizing their sample as a convenience sample, studies taking a phenomenon-based approach to sampling should explicate their sampling principles and describe the process in terms of phenomenon-based sampling.

1. Introduction

This paper addresses a specific type of language sampling in language typology, which I propose to call phenomenon-based sampling.[1] When studying phenomena that are known, on the basis of earlier research, to have a restricted distribution in the languages of the world, it often makes sense to select the sample languages among languages that are already known to exhibit the phenomenon in question, rather than using a sampling method aiming at areally and genealogically balanced world-wide representation of languages. Such phenomena include vowel harmony, egophoricity, switch reference and many others. Naturally, if the research aims include finding out what the world-wide distribution of the phenomenon is, in which areas and families it is found and in which ones it isn't, then a balanced world-wide sample is needed. But if the goals of the study are more in examining aspects of a phenomenon that we already know to be restricted in its areal distribution, then a phenomenon-based sampling method is warranted.

There are many studies that take this type of approach to sampling, so the approach itself is not new, but as far as I know, no closer attention has yet been devoted to it in the sampling literature. The motivation for writing this paper arises mainly from discussions with students working on the typology of phenomena such as the ones mentioned above.[2] So far, existing sampling literature has not given sufficient tools to talk about sampling in such studies. Samples in these kinds of studies are often referred to as convenience samples, but as argued below, the term does not depict the sampling procedure in a proper way. In this paper my aim is to define phenomenon-based sampling and to discuss its main characteristics and its relationship to some

[1] Given that language typology is one of Katia's central research interests, I hope a paper on this topic will fulfil its celebratory function well!

[2] As a result of such discussions, the term phenomenon-based sampling has been mentioned by some students in their theses, e.g. by Keinänen (2025).

other types of sampling in language typology. My aim is not to propose and develop a new method, but rather to identify, compare and discuss some phenomenon-based solutions in earlier literature.

In a strict sense, language sampling refers to well-defined methods designed to ensure that the sample be representative of the population of interest, typically all the world's languages. The term sampling is, however, commonly used in a broader sense to refer to the process of selecting the languages to be included in a typological study, even in cases in which no systematic or explicit principles are used to guide the selection process, which the researcher may simply refer to as convenience sampling. In this paper, I use the term in a broad sense, discussing a wide range of selection processes from real typological studies under the umbrella of sampling, while acknowledging that many of them would not meet the criteria to be called sampling in the strict sense.

The paper is structured as follows. Section 2 provides a background for the discussion of phenomenon-based sampling by introducing some central sampling concepts and by making a brief overview of different types of sampling in language typology. Section 3 constitutes the core of this article, defining phenomenon-based sampling in some more detail and discussing phenomenon-based sampling in prior literature through a survey of relevant articles in the journal *Linguistic Typology* 1997–2024. Section 4 concludes the paper.

2. Sampling and types of samples in typology

Language typology is theoretical linguistics with its empirical basis in large-scale comparison of languages. How to choose the languages that enter into the comparison is a central methodological question in typology. In addition to the word-wide comparative scope of what can be called macro-typology, typological studies may also take a more restricted areal or genealogical perspective, focusing on the languages of a specific area or family (or other genealogical unit), and such studies can be characterized as micro-typology. In this paper, my focus will be in sampling in the macro-typological perspective.

Why do we need to bother about sampling? The problems in selecting languages and the goals of sampling may be different in different studies, but in the most typical macro-typological scenario, we wish to say something general that would be valid for all languages. According to the *Glottolog* (Hammarström & al. 2024), there are more than 8000 languages in the world, and for various reasons, such as the availability of sources and time constraints, it is not possible to include all 8000 in one study. Therefore, we need to work on a subset of the world's languages, i.e. a language sample, and this brings us to the issue of how to choose the languages for the sample in such a way that it serves as a valid empirical basis for the study in question.

Before turning to different types of samples, three central sampling concepts need to be defined: universe, frame and sample. They are familiar from general sampling theory, and they were introduced to typological sampling in Bell 1978, which can be regarded as the first paper devoted to sampling in language typology. The *universe* refers to the population that the study aims to generalize over. In the case of language typology (in the macro sense), the universe is typically the set of all existing languages or even the set of all possible human languages. The *frame* is that part of the universe to which the researcher has access. In the case of language typology, this means the set of languages for which data is available to the typologist. The *sample* is the set of objects selected for observation – in typology the languages for which data is collected and analysed.

The requirements for a sample are different in different studies. In a general macro-typological study, the most important requirement is that the sample be representative of the

universe of all languages that we want to generalize over. To achieve representativity, it is important to pay attention to sample size and to the balanced representation of relevant language groupings, typically areal and genealogical groupings. In other words, areal, genealogical (and possible other biases) should be removed from the sample through a process of stratification. The sampling frame is divided into relevant strata, areal and/or genealogical groupings, from which the sample languages are then picked in a balanced manner.

In linguistic typology, one salient distinction is between *variety samples* and *probability samples* (Rijkhoff & al. 1993, see also Miestamo & al. 2016 for discussion). The goal of variety sampling is to capture as much cross-linguistic variety as possible, whereas probability samples are designed to serve as a basis for statistical testing of universals. To maximize variety, variety samples tend to be very large, and increasing sample size has a positive effect on variety. Stratification is, however, also important: the more balanced a variety sample of a given size is areally/genealogically, the higher the probability that it will capture a high degree of variety. For probability samples, it is even more important that the sample languages be areally and genealogically as independent of each other as possible. Independence of the sampled objects is a prerequisite for many statistical methods used in testing universals. Increasing sample size is not unproblematic for probability sampling – the larger the sample, the more difficult it is to ensure the areal and genealogical independence of the sample languages. Finding the right balance between sample size and independence is a central issue in probability sampling (see Perkins 1989 for discussion). Note however that with recent developments in quantitative typology (see e.g. Bickel 2015), biases may be controlled more efficiently via statistical modelling, using larger datasets without such strict requirements regarding independence, rather than via sampling, which has made probability sampling more or less obsolete in language typology, but variety sampling still remains relevant.

Simple random sampling in language typology would mean taking the list of all languages, e.g. in the *Glottolog*, and picking languages from that list based on randomly generated numbers. Such a procedure would not lead to a balanced sample of the world's languages and has not been used in typological studies. In stratified random sampling, the sampling frame is first divided into areal and/or genealogical groupings (strata) from which the sample languages are then picked randomly. This is a common aim in typological sampling: once the relevant strata have been defined, random selection of languages within these strata is desirable; however, due to the poor availability of adequate sources, pure random selection is very rarely possible.

Quite often typologists refer to their samples as convenience samples.[3] Convenience sampling means that the languages are chosen without any principles or method, but rather picking languages for which data is easily accessible to the researcher. Pure convenience sampling without any method would not count as sampling in the strict sense. Easy access to the researcher can mean, for example, languages whose grammars the researcher has on the bookshelf or that are easily accessible in the institutional library or online. Convenience samples are obviously likely to contain areal and genealogical biases. It should also be noted that the convenience factor may be present in a sample to different degrees. For example, the researcher may aim at a balanced areal-genealogical spread by picking the conveniently available languages is such a way that the sample becomes reasonably balanced, but without using any strict method of stratification.

Another distinction worth mentioning in this context is *a priori* vs. *a posteriori* (or *post hoc*) sampling. *A priori* sampling means that the sampling principles are applied before data collection, i.e. languages are chosen according to a sampling method and data is then collected from the languages of this sample. *A posteriori* sampling means that the sampling principles

[3] The term convenience sampling is used in other (sub)fields as well, see e.g. Buchstaller & Khattab 2013: 76 for how it can be understood in sociolinguistic studies.

are applied after data collection, i.e. data has been collected from a larger set of languages in a less systematic manner and a subset of languages is then taken from this larger set on the basis of an explicit sampling method. Collecting data from a larger set of languages and then only using a smaller subset is of course not very cost-effective, but there may for example be situations where some more specific research questions within a given study need a more restricted sample. Also, sampling data from existing typological databases can be seen as a kind of *a posteriori* sampling.

The purpose of this section has not been to provide an in-depth discussion of types of sampling methods in language typology (see for example Bakker 2011 and Miestamo & al. 2016 for overviews and more discussion). Rather, my aim has been to introduce a few useful concepts and background for the ensuing discussion of a type of sampling that has not been recognized in sampling literature so far, namely phenomenon-based language sampling, to which I turn in Section 3.

3. Phenomenon-based sampling

Phenomenon-based (or phenomenon-driven) sampling is a type of language sampling in which the researcher picks the sample languages from among those languages for which s/he knows, typically on the basis of earlier research, that they exhibit the phenomenon under study. As will be seen below, the knowledge of the existence of the phenomenon in these languages may be acquired in different ways. Vowel harmony, egophoricity and switch reference were mentioned above as phenomena for the study of which a phenomenon-based sampling procedure would in many cases be a good solution. These are phenomena that have a restricted or at least a heavily biased world-wide distribution. Vowel harmony is very common in the languages of northern Eurasia, especially Uralic and Turkic, as well as in large parts of Africa, e.g. in many Bantu languages. Egophoricity is concentrated in areas like the Himalayas and Western China, the Andes and the Caucasus, but absent in most areas of the world. Similarly, switch reference is found in many Papuan and Australian languages, in parts of North and South America and in some languages of Ethiopia, but very rare or absent elsewhere. Further examples of phenomena for which phenomenon-based sampling can be and has been used will be seen in the literature survey below.[4]

To do a typological study of any phenomenon with such a biased distribution, it will, in most cases, be a waste of time and energy to work on a world-wide balanced sample, in which most languages are known beforehand not to exhibit the phenomenon under study. Naturally, if the aims of the study include charting the world-wide distribution of the phenomenon and looking for new languages that might show the phenomenon after all, a balanced world-wide sample is needed. With a phenomenon-based sample, it is always possible that the study will miss languages that are not known by the researcher to show the phenomenon either because the researcher has not encountered it in the literature or because it has not yet been discovered that the language has it. Opting for phenomenon-based sampling, one runs this risk, but in many cases the possible gains from a world-wide sample would be so small and unpredictable that it is not worth the resources one needs to put into it. Depending on the aims of the study and on the expected distribution of the phenomenon, one may also combine phenomenon-based sampling and balanced sampling in one study (more on this below).

[4] It may be noted here that Di Garbo & Napoleão de Souza (2023) call their sampling method phenomenon-based, but their method targets a very different type of phenomenon, namely language contact rather than a strictly linguistic (phonological, morphosyntactic or semantico-pragmatic) phenomenon, and consequently their sampling method is very different from and not directly relevant to what is being discussed in this article.

If the aim of the study is to make generalizations over all languages that exhibit the phenomenon, we can define the sampling universe and sampling frame as follows. The sampling universe is the set of all languages that exhibit the phenomenon irrespective of whether they are known to do so. The sampling frame is the set of all languages known to exhibit the phenomenon, and in more practical terms, the frame is restricted to those languages for which we have access to data. It is among these that the sample is to be selected and there are different solutions to how to do it.

To get a better view of how phenomenon-based sampling has been used in typological research so far, I went through the issues of the journal *Linguistic Typology* (1997–2024), looking for articles that address phenomena that could potentially be approached with a sampling process that could be characterized as phenomenon-based sampling. This journal provides a useful window to recent typological work, but naturally similar solutions have been used in typological studies published in other fora as well, and my survey makes no claims of being exhaustive. The studies that were considered are tentatively grouped into five types of solutions to the choice of languages (numbered I–V below). However, as this was an informal survey with no hard and fast criteria to identify phenomena susceptible for phenomenon-based sampling, and as the grouping is indeed tentative, it seems unnecessary to give any quantitative data of the different types of solutions found. The complete tables of contents of the journal can be retrieved through the journal's website.[5]

I. In a major part of the articles surveyed, languages that exhibit the phenomenon are identified based on existing literature and databases without any systematic sampling procedure. On the other hand, choosing the final sample among these languages may follow systematic criteria. Visser's (2015) study of tensed evidentials identified 36 languages exhibiting the phenomenon, primarily with the help of monographs, edited volumes and individual articles on evdentiality. Some unsystematic searches in newly appeared grammars were also made, but as these did not yield much, this method was not pursued further. Guérin's (2015) study of demonstrative verbs expressing manner examines a sample of 18 languages. The search for these languages started from three languages reported to have the phenomenon in the literature, then looking at languages related to these three and then at grammatical descriptions of other languages easily avaliable to the researcher. In their survey of incorporating verbs, Olthof & al. (2021), use a sample of 50 languages selected from a list of ca. 250 languages (from 82 families) that are reported in the literature to show noun incorporation; the 50 languages were selected so that they would represent different families and that good quality data would be available for them. Mauri and Sansò (2023) examined data from 110 languages in which an associative plural or a similative plural construction is attested, but there is no explicit mention of how these languages were found. Caballero et al. (2008) searched for languages with noun incorporation, using as leads secondary sources and the Autotyp database, as well as their own knowledge of languages, and complemented this with additional surveying of grammars. Their searches yielded a sample of 45 languages with incorporation patterns such that each language comes form a different family or major subfamily. Wu's (2011) study of prenominal relative clauses started from Dryer's (2013) WALS chapter, in which 117 out of 704 langauges show prenominal relatives, and with an additional 58 languages, the sample adds up to 175 languages with prenominal relatives. How the additional 58 languages were found is not explicated in the paper. Lǐ and Ponsford, in their (2018) study of predicative reduplication, state that, following a recommendation by Bybee et al. (1994), their "sampling was led by which languages use reduplication in predication, rather than imposing a prior selection of languages". They end up with a sample of 108 languages with predicative reduplication according to their criteria; however, it is not very clear how the

[5] https://www.degruyter.com/journal/key/lity/html

languages that meet the criteria are found. Vigus (2018) studies antipassive constructions looking at 70 languages found by searching general typological literature, literature on antipassives and reference grammars; there is no explicit mention as to whether reference grammars were searched only for languages for which the literature suggested they would have antipassive constructions or more broadly. Wivell et al. (2024), in their study of ablaut reduplicative structures, constructed a database of 228 languages by looking for descriptions of "reduplication", "word doubling", etc. in grammars that were available to them. From this database a subset of 64 languages that exhibited reduplication with a vowel change in one of the constituents was chosen for further investigation, and this set was further narrowed down to 31 languages according to certain criteria.

II. In a number of the studies in the survey, a broad world-wide language sample is examined and the languages that are found to exhibit the phenomenon in question are selected as the actual sample to be invesigated further. Siewierska's (1998) study of nominal and verbal person marking first examines a 270-language world-wide sample and then selects 157 languages that show the phenomenon (i.e. person marking on both verbs and nouns) for the actual sample to be investigated. Jin and Koening (2021), in their study of expletive negation, go through an unbalanced but very large sample (722 languages) searching the grammars of the sample languages for certain terms, and find 74 languages that exhibit the phenomenon that are then examined in the study. Killian's (2022) study of predicative demonstratives is based on 149 languages that have predicative demonstratives, out of a total sample of 1146 languages examined.[6] In Inglese's (2022) study of middle voice systems, a variety sample of 400 languages is examined, and middle marking is found in 105 languages. An additional set of 24 middle marking languages is added and the total sample size is then 129 languages. Hagège (2008) studies interrogative verbs starting from a sample of 218 languages out of which 28 exhibit the phenomenon and are thus analysed further. It is not explicated how the 218-language sample was constructed, nor how they were surveyed, or how the 28 languages exhibiting the phenomenon were extracted.

III. Sometimes there is no explicit mention of how the languages that are selected for the study are identified, i.e. whether they are found by consulting relevant literature or selected from a larger sample, but they are still listed, so we can refer to them as a sample (in a broad sense). An example of this is Olsson and Hajek's (2003) study of the labial flap. The phenomenon is found relatively widely in Africa and in one Indonesian language. There is no indication of how the phenomenon is found in these languages or whether any other languages have been examined. Even if the phenomenon is heavily concentrated in Africa, the study does have a wider scope, and its goals can be characterized as macro-typological rather than micro-typological; the study is thus relevant in the context of the present discussion.

IV. In some cases, only a few languages that are known to exhibit the phenomenon are discussed. This is the case in the DeLancey's (1997) study of mirativity, which discusses five languages in detail (Turkish, Hare, Sunwar, Lhasa Tibetan, Korean) and mentions a number of other languages more briefly. Similarly, Antonov (2015) discusses verbal allocutivity in five languages in detail (Basque, Pumé, Nambikwara, Mandan, Beja) and also considers data from Japanese and Korean when doing cross-linguistic comparisons. It is dubious whether the sets of languages discussed in these studies can be referred to as language samples, but in any case, as all the languages considered by the researchers are mentioned, the empirical basis of these studies is explicit.

V. Finally, there are some studies where, although a relatively large number of languages is discussed and it is implied that the study is based on a large number of languages, there is no explicit mention of which languages have been looked at and which languages the

[6] In the context of the present volume, it is relevant to mention that, outside the survey of the articles in *Linguistic Typology*, the article by Killian and Gruzdeva (2023) uses the same procedure.

generalizations are based on. This is the case in Aikhenvald's (2015) paper on evidentials and their links with other grammatical categories as well as in Plank's (2005) study of delocutive verbs. In these cases we can hardly talk about sampling.

The first and largest group of studies discussed above represent phenomenon-based sampling in the most prototypical sense: languages exhibiting the phenomenon under study are identified on the basis of earlier research on the phenomenon and the sample languages are then selected from among the languages thus identified. The studies vary as to how the languages that show the phenomenon are identified and how the sample languages are selected among them.

Identifying the languages on the basis of earlier research on the phenomenon, research literature, databases such as WALS or Autotyp, queries directed to experts on relevant e-mail lists, or the researchers' research-based prior knowledge of languages can be called informed identification. If one extends the search to languages that are genealogically or areally close to languages in which the phenomenon is attested in earlier work, as in done e.g. by Guérin (2015), we can still talk about informed identification, as the process is guided by clues from earlier literature. In contrast, identifying languages based on grammar searches without any prior knowledge of whether the phenomenon might be found in the language can be called uninformed identification. The procedure of going through grammars without prior knowledge of the existence of the phenomenon in them is not really different from what is done in variety/probability sampling in general, and thus uninformed identification is not phenomenon-based sampling in the prototypical sense. It may, however, be used to complement phenomenon-based sampling, as is done e.g. by Guérin (2015) and Caballero & al. (2008).[7]

Once the languages exhibiting the phenomenon have been identified and listed, there are different solutions as to how to select the sample languages from this list (which constitutes the sampling frame). The solution to be adopted depends on the number of languages on the list, on resources and on the goals of the study. In some cases, all languages found to exhibit the phenomenon can be included in the sample. But if the list of identified languages is long, it may be useful to sample only a subset of them, following some systematic sampling principles, e.g. to ensure genealogical and/or areal balance. This is done e.g. by Olthof et al. (2021) and Caballero et al. (2008). Stratified sampling within the list of languages exhibiting the phenomenon is done to maximize representativity and variety and similar principles of stratification may be used as in world-wide variety sampling. Note that selecting languages from a list of languages already identified to exhibit the phenomenon might sound like *a posteriori* sampling, but as the researcher has not yet engaged in actual data collection at this point beyond identifying the existence of the phenomenon in those languages, this is not the case.

The second group of studies discussed above use broad world-wide samples at the first stage of the research, going through all languages in those samples without prior knowledge of which languages have the phenomenon, and then select the actual sample from among the languages which exhibit the phenomenon. The total sample used at the first stage is not a phenomenon-based sample and these studies do not use phenomenon-based sampling to avoid going through a lot of languages that do not show the phenomenon. However, the sample selected for closer examination at the second stage can be called a phenomenon-based sample in the sense that the inclusion of languages is based on the existence of the phenomenon. We can thus say that these studies combine broad world-wide sampling and phenomenon-based sampling. This has the obvious advantage that they also manage to chart the world-wide distribution of the phenomenon more systematically, but going through a very large sample

[7] If these studies had listed all the languages for which they consulted grammars, they could perhaps have referred to that list as the total sample increasing thus the explicit cross-linguistic coverage of their survey in a similar way to the studies in the second group.

requires a lot of resources and may not always be needed or may not be possible. The studies in this group vary as to whether the total sample from which the phenomenon-based sample is drawn is based on systematic sampling principles (Inglese 2022) or whether it is a less systematically collected set of languages (Jin & Koenig 2021); the former approach tends to use somewhat smaller samples as stratification enables representativity with smaller sample sizes and thus saves resources. Especially with very large sample sizes, the phenomenon is sometimes harvested from the grammars (semi-)automatically, e.g. by looking for certain terms in electronic versions of the grammars, and only those grammars where the search terms are found will be manually consulted.

It is in connection with the first two groups of studies that we can talk about sampling in any meaningful sense. The studies in the remaining three groups do not follow any explicit principles in how they select the languages to be examined, and there is thus no need to discuss them in any more detail here.

4. Conclusion

In this paper, I have discussed a type of language sampling that I propose to call phenomenon-based, in which, rather than using an extensive sample with a balanced world-wide representation of languages, the sample languages are selected among languages that are already known to exhibit the phenomenon under study. Such a sampling procedure is often warranted when studying linguistic phenomena that have a restricted distribution in the languages of the world. My aim has been to discuss various phenomenon-based solutions found in earlier literature rather than to propose and develop a specific method. I hope that the discussion in this paper can serve as basis for such methodological development in the future.

I wish to conclude the paper by coming briefly back to the term convenience sample. Some studies discussed above refer to their samples as convenience samples, e.g Visser (2015). However, convenience sampling means that languages are selected on the basis of what is convenient and easy for the researcher, e.g. because the grammars of the languages surveyed are easily available to the researcher and no effort is made to apply systematic principles (that might then require further bibliographic searches for grammars not immediately available on the researcher's bookshelf or hard disk). As we have seen above, phenomenon-based sampling is very different from this, as it involves systematic efforts by the researcher, consulting literature and databases and applying specific criteria to identify the languages. Instead of calling this type of sampling with the negatively-connotated, and in my opinion inappropriate, term convenience sampling, we should explicate the sampling principles followed in the study, and adopt the term phenomenon-based sampling to recognize it as a specific type of sampling in its own right.

References

Aikhenvald, A. Y. 2015. Evidentials: Their links with other grammatical categories. *Linguistic Typology* 19 (2). 239–277.

Antonov, A. 2015. Verbal allocutivity in a crosslinguistic perspective. *Linguistic Typology* 19 (1). 55–85.

Bakker, D. 2011. Language sampling. In: J. J. Song (ed.). *The Oxford handbook of linguistic typology*. Oxford: Oxford University Press, 100–127.

Bell, A. 1978. Language samples. In: J. H. Greenberg (ed.). *Universals of human language, Vol. 1: Method & theory*. Stanford: Stanford University Press, 123–156.

Bickel, B. 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In: B. Heine anf H. Narrog (eds.). *The Oxford handbook of linguistic analysis*, 2nd edition. Oxford: Oxford University Press, 901–923.

Buchstaller, I., Khattab, G. 2013. Population samples. In: R. J. Podesva and D. Sharma (eds.). *Research methods in linguistics*. Cambridge: Cambridge University Press, 74–95.

Bybee, J., Perkins, R., Pagliuca W. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago and London: University of Chicago Press.

Caballero, G., Houser, M. J., Marcus, N., McFarland, T., Pycha, A., Toosarvandani, M., Nichols, J. 2008. Nonsyntactic ordering effects in noun incorporation. *Linguistic Typology* 12 (3). 383–421.

DeLancey, S. 1997. Mirativity: The grammatical marking of unexpected Information. *Linguistic Typology* 1 (1). 33–52.

Di Garbo, F., Napoleão de Souza, R. 2023. A sampling technique for worldwide comparisons of language contact scenarios. *Linguistic Typology* 27 (3). 553–589. https://doi.org/10.1515/lingty-2022-0005

Hagège, C. 2008. Towards a typology of interrogative verbs. *Linguistic Typology* 12 (1). 1–44.

Hammarström, H., Forkel, R., Haspelmath, M., Bank, S. 2024. *Glottolog 5.1*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.14006617

Inglese, G. 2022. Towards a typology of middle voice systems. *Linguistic Typology* 26(3). 489–531.

Jin, Y., Koenig J.-P. 2021. A cross-linguistic study of expletive negation. *Linguistic Typology* 25 (1). 39–78.

Keinänen, Satu. 2025. *Functional extensions of evidentials: A cross-linguistic study*. PhD dissertation, University of Helsinki.

Killian, D. 2022. Towards a typology of predicative demonstratives. *Linguistic Typology* 26 (1). 1–41.

Killian, D., Gruzdeva, E. 2023. Towards a typology of demonstrative verbs. *Finnish Journal of Linguistics* 36. 37–92.

Lǐ, Y., Ponsford, D. 2018. Predicative reduplication: Functions, their relationships and iconicities. *Linguistic Typology* 22 (1). 51–117.

Mauri, C., Sansò, A. 2023. Heterogeneous sets: a diachronic typology of associative and similative plurals. *Linguistic Typology* 27 (1). 1–40.

Miestamo, M., Bakker, D., Arppe, A. 2016. Sampling for variety. *Linguistic Typology* 20 (2). 233–296.

Olson, K. S., Hajek, J. 2003. Crosslinguistic insights on the labial flap. *Linguistic Typology* 7 (2). 157–186.

Olthof, M., van Lier, E., Claessen, T., Danielsen, S., Haude, K., Lehmann, N., Mous, M., Verhoeven, E., Visser, E., Vuillermet, M., Wolvengrey, A. 2021. Verb-based restrictions on noun incorporation across languages. *Linguistic Typology* 25 (2). 211–256.

Perkins, R. D. 1989. Statistical techniques for determining language sample size. *Studies in Language* 13. 293–315.

Plank. F. 2005. Delocutive verbs, crosslinguistically. *Linguistic Typology* 9 (3). 459–491.

Rijkhoff, J. Bakker, D., Hengeveld, K., Kahrel, P. 1993. A method of language sampling. *Studies in Language* 17. 169–203.

Siewierska, A. 1998. On nominal and verbal person marking. *Linguistic Typology* 2 (1). 1–55.

Vigus, M. 2018. Antipassive constructions: Correlations of form and function across languages. *Linguistic Typology* 22 (3). 339–384.

Visser, E. 2015. Tensed evidentials: A typological study. *Linguistic Typology* 19 (2). 279–325.

Wivell, G., Miatto, V., Karakaş, A., Kostyszyn, K., Repetti, L. 2024. All about ablaut: a typology of ablaut reduplicative structures. *Linguistic Typology* 28 (3). 505–536.

Wu, T. 2011. The syntax of prenominal relative clauses: A typological study. *Linguistic Typology* 15 (3). 569–623.