



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
Main Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2012

Language description and linguistic typology

Zúñiga, Fernando

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: <https://doi.org/10.5167/uzh-69687>
Book Section
Published Version

Originally published at:
Zúñiga, Fernando (2012). Language description and linguistic typology. In: Ender, Andrea; Leemann, Adrian; Wälchli, B. Methods in contemporary linguistics. Berlin: De Gruyter Mouton, 171-193.

Language description and linguistic typology

Fernando Zúñiga

1. Introduction

The past decade has seen not only a renewed interest in field linguistics and the description of lesser-known and endangered languages, but also the appearance of the more comprehensive undertaking of language documentation as a research field in its own right. Parallel to this, the study of linguistic diversity has noticeably evolved, turning into a complex and sophisticated field. The development of these two intellectual endeavors is mainly due to an increasing awareness of both the severity of language endangerment and the theoretical significance of linguistic diversity, and it has benefited from a remarkable improvement of computing hardware and software, as well as from several simultaneous developments in the worldwide availability and use of information technologies.

The important recent development of these two subfields of linguistics has certainly not gone unnoticed in the literature. When addressing the relationship between them, however, most scholars have concentrated on how and how much typology depends on the data provided by descriptive work, as well as on the usefulness and importance of typologically informed descriptions (cf. e.g. Croft 2003, Epps 2011, and the references therein). Rather than replicating articles that deal with historical issues and questions raised by the results of descriptive and typological enterprises, the present paper focuses on methodological issues raised by their respective objects of study and emphasizes the relevance of some challenges they face. The different sections address the descriptivist's activity (§2), the typologist's job (§3), and some selected challenges on the road ahead for the two subfields and their cooperation (§4).

2. From data collection through language description to language documentation

A well-known article by Pamela Munro defines field linguistics as “the collection of primary linguistic data on the basic grammatical facts of a

relatively little studied language in a natural setting from ordinary speakers, and [...] the analysis and dissemination of such data” (2001: 130). The activities thus delineated are at the core of modern language descriptions, which are customarily understood as data collection plus data description (i.e., linguistic analysis) and meta-description (i.e., information and some analysis of the origin and status of both the data and the description). In turn, language descriptions constitute an indispensable part of language documentation projects. The latter commonly include larger amounts of (written or spoken) texts – often with sociolinguistic and anthropological analyses, or at least annotations and comments – and have to address issues related to copyright, dissemination, and storage.

Instead of concentrating on what distinguishes these three concentric enterprises (i.e. field linguistics, language description, and language documentation), I will concentrate on their core here. I will also assume that the usual and pertinent focus on lesser-known languages does not exclude, say, comparable activities conducted on Mandarin Chinese in Beijing or North American English in Chicago from being considered field-linguistic.

Present-day descriptivists most probably agree on the fact that their activity is directed toward the study of language structure and language use. Nevertheless, even though such an undertaking centers on the representation of data and underlying patterns leading to the production of dictionaries and grammars, current descriptive enterprises differ from their traditional counterparts, especially when framed in documentation projects dealing with lesser-known endangered languages. If they are to be state-of-the-art, they focus on primary data without favoring particular genres or text types, have an explicit concern for both accountability and long-term preservation of the data, and are the product of interdisciplinary teams working in close cooperation with and direct involvement of the speech community (Himmelmann 2006: 15). I will comment in some detail on the focus on primary data in what follows.

Primary data can be gathered from either direct observation or experimental procedures (or experiment-like procedures, like elicitation). Advantages and disadvantages of different primary sources are customarily discussed in courses/textbooks on linguistic field methods; cf. Newman and Ratliff (2001), Gippert, Himmelmann, and Mosel (2006), Crowley (2007), and Bowern (2008). By contrast, secondary data consist of material found in specialized studies and reference grammars. The current consensus in descriptive work is that such sources should not constitute the sole, or principal, foundation on which the description is based.

An issue less frequently addressed than the advantages and shortcomings of questionnaires and corpora, is the fact that “raw” primary data as found in an audio/video recording of a particular situation invariably undergo a process of interpretation before being conveyed to the reader of a traditional language description. Never is such a reader exclusively confronted with unfiltered data presented without any comments or explanations. Even if an author addressing morphosyntactic issues manages to avoid a particularly opaque terminology and framework-specific analytical machinery, the transcription of any single sentence of the object language presupposes some phonological analysis, as well as some analysis of the phonological and grammatical wordhood of the units that constitute the sentence. More often than not, of course, the data are presented via the analytical apparatus chosen by the author of the description in order to represent the alleged fundamental regularities of linguistic structure. It is worth mentioning that the situation is somewhat different with recent comprehensive language documentation projects, which sometimes do include the presentation of raw primary data in different audio/video formats.

By a different token, the analyst must make numerous decisions concerning the status granted to the particular situation in which language has been used. Many of these choices may be explicit (e.g. “this is colloquial language / allegro speech,” “such an utterance is felicitous under the following circumstances,” “most speakers agree on these grammaticality judgments”), but the readers of the description have to rely on the descriptivist’s judgments, i.e. his/her interpretive filters. Available descriptions differ greatly as to the amount of meta-information provided; therefore, they also differ as to how reliable they are as sources of material that is readily comparable across languages.

Finally, the focus on primary data is not the only thing related to data sources that has changed over the last decades. Wälchli (2007) and Epps (2011) rightly point out that intralinguistic structural variation has tended to be underrepresented in traditional descriptions, which have favored normalized representations of typical patterns and neglected unsystematic or infrequent structures. (This is related to the bias toward particular genres and texts found in less recent descriptions, especially those of languages that have a written tradition. Narrative written texts produced by arguably influential male adults have tended to be overrepresented, for example, as data source for the traditional descriptions of classical languages.) Recent studies emphasize the need for less restrictive data collection techniques that allow “to support claimed generalizations with multiple empirical sources

of converging evidence, including observations of ecologically natural language use” (Bresnan 2007: 302).

3. Linguistic typology

The systematic study of crosslinguistic variation is not a new field, but the way typology is conceived of and practiced has changed markedly since its beginnings. The founders of the discipline in the early 19th century, viz. the Schlegel brothers and Wilhelm von Humboldt, were primarily concerned with the morphological classification of languages.¹ Greenberg (1963) and his many followers were also concerned with phonetics, phonology, and syntax (particularly, but not exclusively, constituent order), and most modern practitioners have gradually moved away from so-called holistic classifications toward the study of partial subsystems of language. While early thinkers like Steinthal, Finck, Lewy, and Mathesius conceptualized crosslinguistic classification as formally and causally connected to characterology – note that the French 17th-century interest in *le génie de la langue* is famously present as “basic plan” or “structural ‘genius’” in Sapir’s influential work (1921) – , 20th-century typology started focusing on individual domains of language structure.² In recent decades, the emphasis has also shifted from attempting to identify absolute universals (i.e. statements that hold true for all languages without exceptions) to discovering and explaining statistical universals (i.e. statements that hold true for many languages). Lastly, typology has traditionally concentrated on morphosyntactic as well as phonetic/phonological diversity and has ventured into the systematic exploration of lexical patterns comparatively recently; see Brown (2001), Koch (2001), Koptjevskaja-Tamm, Vanhove, and Koch (2007), and Haspelmath and Tadmor (2009) for some discussion and Evans (2011) for a more general picture.

Even though a primary concern of present-day typology is the categorization of linguistic phenomena into types, some scholars think of the present and the future of the discipline in terms that are less restrictive than such a textbook definition would suggest. Nichols (2007: 236), for instance, says that

what we call typology is not properly a subfield of linguistics but is simply framework-neutral analysis and theory plus some of the common applications of such analysis (which include crosslinguistic comparison, geographical mapping, cladistics, and reconstruction).

Similarly, Croft (2007: 80) says he follows Greenberg in considering typology “an empirical, i.e. scientific, investigation into the nature of language. [...] Typology is a theoretical approach to language, with increasingly well-established methods and results.” Lazard (2005: 1–2) pertinently quotes Hjelmslev (1970: 96) with respect to the importance of typological undertakings: “An exhaustive linguistic typology is, in fact, the biggest and most important task facing linguistics. [...] Only through typology does linguistics rise to quite general viewpoints and become a science.”

In this light, it is perhaps unsurprising that Croft (2003: 1–2) presents the following threefold categorization of typological inquiries: what is at the center of attention are classifications, generalizations, and explanations. Typological classification is concerned with structural types (formerly of languages, nowadays of small-scale phenomena, as mentioned above). The second kind of typology examines systematic crosslinguistic patterns, and the third is an “approach to linguistic theorizing, or more precisely a methodology of linguistic analysis that gives rise to different kinds of linguistic theories than found in other approaches [such as American structuralism and generative grammar, FZ].” While these three kinds of typology can be thought of as stages in the development of the discipline (roughly: early, Greenbergian, and modern), they also correspond to partial stages of any empirical scientific analysis, viz. the observation and classification of phenomena, the generalization over the observations and classifications, and the explanation of the generalizations.

Finally note that Bickel’s (2007: 248) characterization of typology is threefold like Croft’s, but it reflects a noteworthy recent development path of the discipline:

Modern typology is a discipline that develops variables for capturing similarities and differences of structures both within and across languages (qualitative typology), explores clusters and skewings in the distribution of these variables (quantitative typology), and proposes theories that explain the clusters and skewings (theoretical typology).

Such a view casts the net more widely in at least two directions: typology is now explicitly seen as studying both crosslinguistic and intralinguistic variation, and the study of “generalizations over observations” is not limited to the study of linguistic phenomena but must include elements studied by disciplines concerned with geography, history, sociology, etc. and the use of quantitative methods of inquiry. Crucially, Bickel (2007: 239) suggests that

typology has begun to emancipate itself from [one of the same goals as generative grammar, viz. to determine the limits of possible human languages and, thereby, to contribute to a universal theory of grammar] and to turn from a method into a full-fledged discipline, with its own research agenda, its own theories, its own problems.

4. Objects of study, methods, and challenges

This section addresses a number of questions related to current language description and linguistic typology with respect to their objects of study, the methods they employ, and some challenges they face. Rather than giving an exhaustive treatment to these far-reaching domains, I will limit myself to mentioning some selected issues raised by what Epps (2011) has called the “continuing partnership” between documentation and typology (§4.1), as well as some problems that arise in the context of the three fundamental methodological domains mentioned by Croft (2003: 8–30), viz. sampling (§4.2), data sources (§4.3), and crosslinguistic comparison (§4.4 and §4.5). Finally, Subsection 4.6 points out two further challenges faced by typological studies: taking intralinguistic variation into account and explicitly incorporating variation as something in need of explanation.

4.1. Some basic challenges faced by the “continuing partnership” between documentation and typology

Theoreticians and practitioners regard language documentation and linguistic typology as organically related in several ways. First, the two subfields have benefited from each other; not only is Epps’s discussion of Hup D[ifferential] O[bject] M[arking] in her descriptive grammar, for instance, informed by the typological literature on the topic, but it also contributes to a more refined typology via its account of the role played by nominal number marking (Epps 2008: 170f, 2009). Second, “[i]t is documentary linguistics that gives typologists access to these usage-based data; at the same time, typological interest in such diverse phenomena highlights the need for documentation to be thorough, broadly inclusive, and ethnographically rich” (Epps 2011: 642). Lastly, with respect to the current state of both subdisciplines, Epps says that

[t]he goal of refining our typological focus is well served by contemporary documentary methodology, which stresses collection of a large and diverse

corpus. [...] [D]ata collection and analysis must focus on language *in use*, and takes the perspective that “linguistic meaning cannot be treated separately from the ‘encyclopedic’ content of the relevant culture and society” (Hudson 2007: 7). [...] Such an approach is essential if we are to achieve a more complex typological understanding of diversity. (Epps 2011: 639; emphasis in the original)

Thus, descriptivists and typologists conceive their fields as fundamentally entwined: hardly anyone would deny that poor descriptions seriously compromise typological investigations based on them, and that shaky typologies at best fail to inform, and at worst misguide, the recording of individual language structures. This poses an evident challenge for authors, publishers, librarians, and archivists: not only language documentation materials but also typological studies (e.g. Shopen 2007) must be as solid, and as widely available, as possible.

There is reasonable consensus on what counts as good descriptive material: the description must meet the dual challenge of enabling and facilitating crosslinguistic comparison while remaining “true to the languages themselves, without forcing them into ill-fitting predetermined categories” (Epps 2011: 648). Similarly, it is generally acknowledged that sound typological research must not oversimplify the “intricacies and complexities” that characterize particular linguistic structures. In practice, of course, this may not be an easy task. Descriptive grammars are multi-purpose artifacts, and they are used by theoretical linguists of all persuasions, by typologists, and by people involved in language revitalization and teaching. Moreover, descriptive work is usually consulted by scholars belonging to different generations – unlike many typological studies (and, to some extent, other products of language documentation), which can be rather short-lived.

Let me conclude this subsection by noting that it is possible to investigate the continuing partnership between typology and description by (i) conceiving the three concentric dimensions of the former mentioned in §3 as related in various ways to the three concentric enterprises of the latter introduced in §2, and (ii) studying how these dimensions interact and inform one another. Typology proceeds from variables and classification through the analysis of clusters and generalizations to the explanation and construction of theories. In turn, documentation proceeds from data collection and storage through data description and meta-description to the analysis of the documentation as a whole. Against this background and in addition to the example of Hup DOM given above, Chapter 7 of Joseph’s descriptive grammar of Rabha (2007) can be used to illustrate the most

straightforward feed-back between classification and description and what some comprehensive present-day grammars look like: the 170-page-long chapter is a “correlative analysis” of Rabha and two other closely related Tibeto-Burman languages, viz. Bodo and Garo. The phonology, the noun classes and the noun morphology, and the verb morphology of these languages are subjected to detailed comparative scrutiny there, which in turn yields illuminating insights for their diachronic, synchronic, and typological characterization.³

4.2. Data sources

The issue of data sources employed in typological inquiries is related to the question of data sources used for language documentation, which was mentioned in Section 2 above. Croft concludes that “[n]o source of data – native consultants, actual texts or descriptive grammars – is perfect; but any and all sources can provide relevant data when used judiciously” (2007: 30). He aptly points out some of the shortcomings of such sources; data via traditional elicitation techniques do not need to accurately represent actual language use, the design and the application of good elicitation questionnaires are difficult, few available texts include face-to-face spoken conversation, and descriptive grammars show biases of different kinds and even gaps, but are better than secondary sources.

An example from my own descriptive work can illustrate some difficulties that have been largely neglected until recently; other practitioners will have no trouble recognizing analogous or related phenomena in their work. Mapudungun is an Amerindian language spoken in Chile and Argentina, and has been in contact with the local varieties of Spanish for several centuries. Both Mapudungun and Spanish show DOM, i.e., nonagentive arguments of transitive clauses are marked differently depending on some of their properties. To be sure, there are some structural differences to be noted; Mapudungun has no case or adpositional marking of core syntactic arguments comparable to the Romance functional equivalents, and DOM appears as a verbal suffix *-fi* alternating not only with a null marker in direct verb forms but also with nonzero markers in inverse verb forms; Spanish DOM, by contrast, appears formally as the opposition between the preposition *a* and zero:

(1) Mapudungun and Spanish DOM (p.k.)

- a. *Pe-fi-n* *Juan.* / *Vi* **(a) Juan.*
 see-3O₁-1SG.IND J. saw[.1SG] to J.
 'I saw Juan.'
- b. *Pe-Ø-n* *kura.* / *Vi* **(a) la piedra.*
 see-3O₂-1SG.IND stone saw[.1SG] to ART stone
 'I saw the stone.'
- c. *Pe-e-n-mew* *Juan.* / *Juan me vio.*
 see-INV-1SG.IND-3A J. J. 1SG.O saw.3SG
 'Juan saw me.'

What is important here is that the conditions under which DOM is found appear to be quite different in both languages. The examples in (1) suggest that objects with human referents trigger the differential marking while those with nonhuman ones do not, and that the conditions governing the direct/inverse opposition in Mapudungun account for the additional difference between the structures. Nevertheless, factors related to the pragmatic status of the arguments and to discourse structure are of paramount importance in Mapudungun, whereas animacy and specificity seem to be the main factors governing Spanish DOM (even though some lexical issues make the Spanish picture more complicated and semantic and referential factors do play a role in Mapudungun as well; cf. Zúñiga 2010 and references therein for details.) Roughly, specific human third persons trigger DOM and inversion when they are especially salient objects and subjects respectively, but typically only at key inflection points in the text, where these referents are to be (re)activated.⁴

The exploration of how Mapudungun DOM works is relevant for the assessment of data sources on a number of grounds. Older grammatical descriptions of Mapudungun pay almost no attention to pragmatic and discourse factors. Everyday conversation shows only sporadic occurrences of *fi*-marked verbs and inverse forms corresponding to 3↔3 interactions, whereas traditional narratives like the *epew* (in which numerous 3↔3 verb forms appear, some of which are instances of DOM or inversion) are relatively difficult to elicit.⁵ Questionnaires and the direct, possibly even experiment-prompted, elicitation of isolated clauses conducted in Spanish, even in carefully provided contexts, will seldom be useful, since the danger of obtaining a distorted use of DOM forms is rather high (either too many under the influence of the Spanish prompt or too few due to hypercorrection). Elicitation conducted in Mapudungun will probably also miss the fact

that the decisive factors triggering the alternations are related to discourse factors visible only in longer texts.

A better understanding of Mapudungun DOM leads, among many other things and together with relevant evidence from other languages (cf. Zúñiga 2007, Iemmolo 2010, and Dalrymple and Nikolaeva 2011), to a refined typology of case marking, agreement, and alignment that integrates not only properties of (referents of) NPs and clause-level phenomena but also discourse-level or at least text-level phenomena. Thus, the emphasis on useful primary data in this context is of paramount importance for the adequate description of numerous lesser-known languages, a rethinking of familiar accounts of DOM in languages like Spanish, Turkish, or Persian (where the connection to discourse factors has been occasionally noted, cf. chapter 6 in Comrie 1981), and a better typology of grammatical relations.

4.3. Sampling

Quantitative, especially statistical, methods have become a widely used tool in present-day typology. It should not come as a surprise, therefore, that the discipline has to solve the nontrivial sampling problem, which is indirectly related to language description via language classification. In order to minimize several kinds of bias (genetic, areal, typological, and cultural), typologists employ sampling techniques adapted from those used in other disciplines; see Bakker (2011) and the references therein for details and discussion. A much more interesting dimension of the sampling problem is *directly* related to language description via the definition of the unit of study – an issue that typology has begun addressing in a principled way only relatively recently. For instance, work by Bickel (2011) suggests that, for some typological questions related to referential density and case patterns, it may be more relevant to look at units that are either higher or lower than the dialect/language unit (i.e. language groups within a family or clusters of particular idiolectal registers). The twofold question of how to best sample what kinds of linguistic units is likely to be addressed by significant and illuminating studies in the near future.

4.4. Crosslinguistic comparison I: Basics

The issue of crosslinguistic comparability has a long-standing history in the humanities and has received renewed attention in recent typological studies. Perhaps somewhat expectedly, what is at the center of the current de-

bate is *how* crosslinguistic comparison should best proceed, rather than *whether* it is feasible at all.

According to Croft, for example, “the variation in structure makes it impossible to use structural criteria, or only structural criteria, to identify grammatical categories across languages” (2003: 13), and the solution lies in the use of “external” (i.e. semantic and pragmatic) definitions. In Croft’s view, the “standard research strategy” in typology can be described as follows. First, the analyst determines the particular semantico-pragmatic structure or situation type to be studied. Second, s/he examines the morphosyntactic strategies employed to encode that situation type in different languages. Finally, s/he looks for dependencies between the constructions found and other linguistic factors (“other structural features, other external functions expressed by the construction in question, or both”) (Croft 2003: 14). Croft also acknowledges the validity of “derived” structural definitions for the constructions in question (i.e. those including semantic-pragmatic components and morphosyntactic elements); the choice between them and purely external definitions is determined by the purpose of the study. For example, the appropriate definition of the subjunctive is purely external for an exploration of the realm of modality but should be derived if the phenomena to be covered are the different structures of complex sentences (Croft 2003: 18).

At first sight, Lazard (2005) might seem to disagree with Croft by observing that both structural criteria and “semantic substance” are language-specific. He follows the Saussurean view of thought as “amorphous” before it is structured into (language-)specific signifier/signified correspondences and proposes to resort to a particular type of “intuition”: the analyst must make so-called arbitrary conceptual frameworks (ACFs) the point of departure for typological research. These consist of explicit definitions and/or propositions about linguistic and/or extralinguistic phenomena, are arbitrary (i.e. freely chosen by the analyst) but ideally informed by “a wide experience of different languages.” They preferably concern limited domains of grammatical systems or lexical fields, are research tools rather than empirically falsifiable hypotheses, and are provisional (i.e., they should be replaced by other ACFs if they do not lead to interesting discoveries) (Lazard 2005: 8).

Nevertheless, Lazard’s (2005: 16) discussion of grammatical voice makes it clear that he regards the procedure employed by Croft (2001: 283-319) when charting the “conceptual space” of that category as sound:

The author thoroughly investigates the morphosyntactic data, brings them together with a well defined conceptual space, and arrives at the discovery of an accurately formulated invariant, which consists of a constant relationship emerging from the diverse shapes of the correlation between structure and function or, in another terminology, between signifiant and signifié.

Besides the still somewhat controversial point concerning the nature of universals (“invariants are neither forms nor meanings; they are relationships appearing in the correlation between forms and meanings,” p. 16), it is important to observe that Croft uses, in addition to morphosyntactic concepts, the notions of agent and patient, which are not considered to be language-specific and are therefore viable for the exploration of phenomena like (anti)passivization, (anti)causativization, and applicativization.

The existence and the status of the *tertium comparationis* have been addressed in more detail by Haspelmath in a series of recent studies (2007, 2009, 2010). Contrary to what is customarily taken for granted in Chomskyan crosslinguistic work, Haspelmath claims that pre-established structural categories of grammar do not exist: it is the job of neither language description nor linguistic typology to attempt to create, motivate, or discover such entities. Rather, “[I]anguage describers have to create language-particular structural categories for their language” (2007: 125); with respect to the notions used in typology, morphosyntactic comparison must be “semantically based” (p. 126). Instead of allegedly spurious crosslinguistic categories, typology must (and customarily does) employ “comparative concepts,” as detailed in Haspelmath (2009). These are specifically designed for comparison purposes and defined via concepts potentially applicable to any human language, viz. conceptual-semantic concepts, general formal concepts, and other (more primitive) comparative concepts (Haspelmath 2010).

Rather than discussing the definition and the nature of the best *tertium comparationis* in detail here, I want to highlight the fact that such fundamental questions are currently being debated. While some scholars argue that language-individual description and crosslinguistic comparison use related but different sets of concepts (e.g. Lazard 2006 and the references given above of work by Haspelmath), other typologists contend that

typological survey is identical to language-specific analysis and consists in detailed description of properties. To the extent that we learn more about language specific properties, the better can we define fine-grained typological variables. And such fine-grained variables provide in return the ideal questionnaires for fieldwork. (Bickel 2010: 93)

Such a view naturally leads to a direct connection between descriptive data and typological data (i.e. to something different from a derivative relation between them, like in Haspelmath's comparative concepts, as well as different from a sophisticated isomorphism between data/concepts belonging to different realms, like in Seiler's UNITYP framework, cf. Seiler 2001). Several typological databases actually operate in this fashion, linking descriptive and typological data/concepts in a straightforward way (e.g. the Leiden Stresstyp project on metrical systems and the Berkeley/Zurich Autotyp project on morphosyntax and word domains).

4.5. Crosslinguistic comparison II: Some challenges

Bickel (2007) pertinently emphasizes that typology has to find ways to capture, measure, and elucidate linguistic diversity rather than reducing it and thereby potentially explaining it away. Systematic variation across languages can indeed become overwhelming, and much of current scholarship in the field is devoted to developing methods that enable us to adequately deal with crosslinguistic diversity. To my knowledge, however, even though linguists habitually acknowledge the importance of systematic (and less systematic) variation, typologists have just started to deal with the issue in a principled way.

In this context, let me comment on the comparison of (pro)nominal paradigms across languages. The case systems of Slavonic, Indo-Aryan, and Kartvelian languages, for instance, are often compared in run-of-the-mill typological studies, but the parameters at the center of attention are usually semantic and syntactic. The relevance of questions related to pragmatics (and, more widely put, text and discourse structure) has been incorporated into the analysis comparatively recently,⁶ and aspects of case morphology that touch upon sociolinguistics and other cultural factors have been neglected unless they are prominent, like in Samoan.⁷

Even in the absence of such factors, comparability is not always as mechanical as textbook introductions may suggest. Person-based ergativity splits, for example, can be compared across languages by calculating an index that quantifies the degree of ergativity. (This is the kind of question, or rather one kind of analytic tool, some quantitative studies currently gravitate towards.⁸) Standard Basque first and second person pronouns, for instance, show an ergativity index of 80% (=4/5), since four out of five items distinguish an absolutive (S/O) and an ergative (A) form (Table 1).

Table 1. Selected Basque pronouns (p.k.)

	O	S	A
1SG	<i>ni</i>		<i>nik</i>
2SG ₁	<i>hi</i>		<i>hik</i>
1PL	<i>gu</i>		<i>guk</i>
2SG ₂	<i>zu</i>		<i>zuk</i>
2PL		<i>zuek</i>	

By contrast, Dyrirbal shows an ergativity index of 0% (=0/6) for these pronouns, since all items pattern accusatively (Table 2).

Table 2. Selected Dyrirbal pronouns (Dixon 1972: 50f.)

	O	S	A
1SG	<i>ɲayguna</i>		<i>ɲadya</i>
1DU	<i>ɲalidyina</i>		<i>ɲalidyi</i>
1PL	<i>ɲanadyina</i>		<i>ɲanadyi</i>
2SG	<i>ɲinuna</i>		<i>ɲinda</i>
2DU	<i>ɲubaladyina</i>		<i>ɲubaladyi</i>
2PL	<i>ɲuradyina</i>		<i>ɲuradyi</i>

Note that an adequate understanding of what such an ergativity index represents may be more elusive than it appears at first glance. For example, paradigms may not be equally complex: Basque shows two 2SG forms but lacks the specific dual forms found in Dyrirbal, so actually more items could align in different ways in the latter language than in Basque. But even if both paradigms had the same number of *bona fide* items, measuring the degree of ergativity of both paradigms would be computationally simple but conceptually nontrivial. What are the exact limits of the paradigm (and, somewhat more vexingly, can the analyst afford not to postulate them)? Many grammatical descriptions of Basque include the third person items *hura* (ABS.SG), *hark* (ERG.SG), and *haiek* (PL) in some versions of the pronominal paradigm (in which case the complete paradigm would have an ergativity index of $5/7 = 71.43\%$), but these elements are actually distal demonstratives. If all three demonstratives were included – they all pattern like the distal item – , the ergativity index would be $7/11 = 63.63\%$. Similarly, Dyrirbal has the absolutive noun class markers *bayi*, *balan*, *balam*, and *bala*, as well as their ergative counterparts, and from Dixon’s description it is clear that they are both like and unlike first and second person

pronouns in several ways, rendering their inclusion in the paradigm as third person pronouns problematic. (There are additional elements in paradigmatic opposition to these noun class markers, which differ from them along the dimensions of location and visibility; cf. Dixon 1972: 45). Lastly, recall that the Silverstein hierarchy used to make sense of such ergativity splits includes nouns, which pattern like the third person markers in Dyirbal but unlike the demonstratives in Basque, since in the latter language nouns distinguish absolutive from ergative forms not only in the singular but also in the plural. As a consequence, the use to which such an ergativity index is put needs to be discussed in detail if the tool is to be helpful and, possibly, powerful.

Such interpretive problems are not limited to demonstratives and noun class markers and cast some doubt on the usefulness of the simple version of the index presented at the outset of this discussion. The Basque second person singular pronoun *hi(k)* is actually unlike its default counterpart *zu(k)* in that the former is not only obsolescent in some varieties of the language but also heavily restricted on social and pragmatic grounds (as are the allocutive verb forms that mark gender of the *hi(k)*-addressee even if s/he is not a semantic argument of the verb). Roughly, elderly men in rural areas talking among themselves who are close to each other use them most often, while young women in urban areas who are strangers use them most seldom (Amorrortu 2003: 144f). Similarly, is French *on* ‘one, we’ to be considered as well as *je* ‘I’, *tu* ‘you (SG)’, etc. for purposes of comparing case/agreement phenomena? (In the terms phrased by Munro in the definition quoted at the beginning of Section 2, just what are the “basic grammatical facts” of French for the purpose of crosslinguistic comparison here?) More generally, are imperfectly grammaticalized demonstratives or lexical NPs part of the paradigm? Supposing we can reach a reasonable consensus on how to delimit the paradigms in the languages under study, how should we weigh the different forms that constitute them: based on their frequency, on their degree of grammaticalization, or evenly?

For some varieties of Brazilian Portuguese, for example, the latter option would mean that obsolescent *tu* ‘you (SG)’ and robust *ele* ‘he’ are treated on a par, which has advantages for some typological questions but possibly disadvantages for others, like those related to language change. A similar argument applies to the varied and fairly complex use of *voseo* forms in both pronominal and verbal paradigms of numerous varieties of Central and South American Spanish – all of which raises the issue of meaningfully delimiting the particular language varieties, including regis-

ters, under study. For example, the pronoun *vos* ‘you (SG)’ is more widely used in the varieties of Spanish spoken in Buenos Aires and surrounding areas than in those of Santiago de Chile or Caracas (i.e., it is used by speakers of nearly all sociolects in numerous contexts where the latter varieties have *tú* instead). This would arguably justify including *vos* in the River Plate Spanish paradigm, perhaps even at the expense of *tú*, while its status in the Chilean and Venezuelan varieties would be more clearly debatable.⁹ In Munro’s terms, this amounts to a thorough and comprehensive account of the “natural setting” in which the language is used by “ordinary speakers,” including questioning what counts as natural setting, ordinary speaker, and even “the language.”

I do not want to suggest that the way descriptive studies and morpho-syntactic typology have dealt with person, number, gender, case, and agreement in the recent past is useless or fundamentally flawed. Rather, my point is simply that it is time for descriptivists to document and discuss in greater detail what has been excluded and/or neglected in order to arrive at the construct called “language X.” And even more importantly, it is time for typologists to incorporate into their accounts a number of well-known linguistic phenomena hitherto abstracted from. The availability of ever-improving technical resources and the conceptual development of the fields make such steps not only possible but also welcome.

4.6. Further challenges for linguistic typology

As stated repeatedly above, not all challenges to linguistic typology come from issues related to crosslinguistic comparability. Intralinguistic diversity, which has long been neglected in both description and typology, is equally important. Although current documentation projects can help remedy this situation by being less restrictive than their predecessors, linguistic studies face a nontrivial threefold challenge in this respect. First, it is not enough to document as much within-language variation as possible in current and future projects; within-language variation also needs to be studied in endangered languages that have already been described. Second, and more fundamentally, descriptive and theoretical linguists have to refine our present understanding of intralinguistic variation phenomena based on both well-known and lesser-known languages. Third, typologists must address the question of comparability not only of patterns but also of variation of patterns.

The latter challenge is a more complex side of the diversity problem, and it means that typology needs to venture into deeper waters in order to adequately ascertain and explain the nature of intralinguistic and crosslinguistic variation. Both descriptivists and typologists are aware of the significant variation languages can display; the challenge consists in contributing to our knowledge of such variation not only by documenting it but also by integrating it into our accounts of language diversity as what it is: something in need of explanation. In other words, not only do we need to introduce relevant phenomena that are currently missing from our models of linguistic diversity as *independent* variables; we also need to entertain the possibility that they are sometimes best treated as *dependent* variables. To ask why some pronouns, ergative markers, or inverse clauses are pragmatically or sociolinguistically conditioned while others are not is arguably more than a mere *jeu d'esprit*, but linguistic typology is – to my knowledge – not yet in a position to give a principled answer to such questions.

In addition to the strictly linguistic challenges this poses, it also leads to recognizing that not only field linguistics, language description, and language documentation need to take into account the relevant paralinguistic and extralinguistic dimensions alluded in the quote at the beginning of Subsection 4.1 in order to do a better job: typology also has to find realistic but illuminating ways not to explain explaining factors and explananda away. To quote from Epps's lucid article one last time:

[Why diversity is the way it is] demands the consideration of multiple variables: not only universal preferences, but also geographical and (genetic) genealogical distributions, diachronic change, and the interaction between language and social, cognitive, and cultural factors. (Epps 2011: 640)

Acknowledgements

I am indebted to two anonymous reviewers for their valuable comments on a previous version of this article. Not only did some of these proposals help me improve the paper in terms of its readability, but several of the reviewers' useful suggestions also concerned argumentation and content. I would also like to thank Bernhard Wälchli for taking his editorial duties so seriously. Needless to say, all errors, misconceptions, and misrepresentations remain my own.

Notes

1. It is seldom mentioned in surveys that there were some forerunners whose endeavors pointed in the direction of modern typology, viz. Tommaso Campanella (1568-1639), Gabriel Girard (1677-1748), and Nicolas Beauzée (1717-1789). The proposal found in Adam Smith's *Dissertation on the Origin of Languages* (1761) can also be regarded as close in spirit (but arguably not in letter) to the Schlegel-Humboldt typology. See Bosson (2001) for details.
2. Some of Coseriu's work (1988a, 1988b, 1990) exemplifies and discusses this transition; see Aschenberg (2001) for details. See also Schmid (this volume) for an interesting discussion of the holistic/partial issue in the context of phonological typology.
3. I am indebted to an anonymous reviewer for helping me better present the relationship between the two concentric inquiries and its relevance.
4. First and second persons are different in that they invariably trigger DOM and inversion. Such asymmetries between speech act participants and third persons are well known from both descriptive and typological studies.
5. In most Chilean Mapuche communities I am acquainted with, it is elderly male speakers of some authority that are entitled to tell these somewhat conventionalized *epew* – at least in their longer versions –, typically in winter nights, around the hearth, and to familiar audiences. Unfamiliar non-Mapuche researchers wielding recording equipment may or may not be regarded as intrusive in other contexts, but they are not likely to receive permission to fully document such narratives unless they have successfully worked closely with particular individuals in specific communities over an extended period of time (usually, several years).
6. A recent example is Valenzuela (2011), which characterizes the use of the ergative enclitic in Shiwilu and Shawi (Kawapanan; Peru) as crucially linked to discourse factors. A similar phenomenon is found in Zaparoan, and related phenomena have been reported in Chibchan, Arawakan, Tibeto-Burman, and Australian languages (cf. Valenzuela 2011: 116 and her relevant references).
7. Roughly, speakers of Samoan use the ergative case much less frequently in the complex social event called *fono* than in everyday communication. See Duranti (1994: 85f, 125f) for details.
8. Such quantitative assessment of the degree of ergativity for typological purposes has been proposed in recent work conducted by (formerly) Leipzig-based typologists (cf. e.g. Witzlack-Makarevich et al. 2010 and Bickel et al. 2010).
9. The question of verbal *voseo* is analogous but somewhat more complicated. *Voseo* forms are prominent and robust in River Plate Spanish in the indicative and imperative but rather stigmatized and arguably more marginal in the subjunctive. In Chilean verbal *voseo*, by contrast, there are no imperative forms, the indicative and the subjunctive are equally robust, and there is no asymmetric stigma to them. How should the analyst compare such verbal paradigms in

terms of structure, and how much of actual use should be taken into account at what stage of the analysis? Rather than suggesting that such questions do not have (satisfactory) answers, my point here is simply that both descriptivists and typologists must try to give good answers to them.

References

- Amorrortu, Estibaliz
2003 *Basque Sociolinguistics: Language, Society and Culture*. Reno: Center for Basque Studies, University of Nevada.
- Aschenberg, Heidi
2001 Typologie als Charakterologie. In *Language Typology and Language Universals. An International Handbook, Volume 1*, Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), 266–274. Berlin: Walter de Gruyter.
- Bakker, Dik
2011 Language sampling. In *The Oxford Handbook of Linguistic Typology*, Jae Jung Song (ed.), 100–127. Oxford: Oxford University Press.
- Bickel, Balthasar
2007 Typology in the 21st century: major current developments. *Linguistic Typology* 11 (1): 239–251.
2010 Capturing particulars and universals in clause linkage. A multivariate approach. In *Clause Linkage and Clause Hierarchy: Syntax and Pragmatics*, Isabelle Bril (ed.), 51–101. Amsterdam: John Benjamins.
2011 Putting variation center stage: beyond ‘language’ (or ‘dialect’) as the basic data unit in typology (and elsewhere). Keynote address given at the conference ‘Variation and Typology: New trends in Syntactic Research,’ Helsinki, August 25.
- Bickel, Balthasar, Alena Witzlack-Makarevich, and Taras Zakharko
2010 Case alignment across the lexicon. Paper read at the conference ‘Syntax of the World’s Languages IV,’ Lyon, September 23–26.
- Bossong, Georg
2001 Die Anfänge typologischen Denkens im europäischen Rationalismus. In *Language Typology and Language Universals. An International Handbook, Volume 1*, Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), 249–264. Berlin: Walter de Gruyter.
- Bowern, Claire
2008 *Linguistic Fieldwork: A Practical Guide*. London: Palgrave Macmillan.

- Bresnan, Joan
 2007 A few lessons from typology. *Linguistic Typology* 11 (1): 297–306.
- Brown, Cecil
 2001 Lexical typology from an anthropological point of view. In *Language Typology and Language Universals. An International Handbook, Volume 2*, Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), 1178–1190. Berlin: Walter de Gruyter.
- Comrie, Bernard
 1981 *Language Universals and Linguistic Typology: Syntax and Morphology*. Oxford: Blackwell / Chicago: University of Chicago Press.
- Coseriu, Eugenio
 1988a Der Sinn der Sprachtypologie. In *Energieia und Ergon. Sprachliche Variation — Sprachgeschichte — Sprachtypologie, Vol. 1*, Jörn Albrecht, Jens Lütcke, and Harald Thun (eds.), 161–172. Tübingen: Gunter Narr.
 1988b Sprachtypologie und Typologie von sprachlichen Verfahren. In *Energieia und Ergon. Sprachliche Variation — Sprachgeschichte — Sprachtypologie, Vol. 1*, Jörn Albrecht, Jens Lütcke, and Harald Thun (eds.), 195–206. Tübingen: Gunter Narr.
 1990 Überblick über die Ergebnisse der Plenarsitzung ‘Typologie: ganzheitliche Typologie versus Teiltypologie’. In *Proceedings of the 14th International Congress of Linguists I*, Werner Bahner, Joachim Schildt, and Dieter Viehweger (eds.), 237–242. Berlin: Akademie Verlag.
- Croft, William
 2001 *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
 2003 *Typology and Universals*. Cambridge: Cambridge University Press.
 2007 Typology and linguistic theory in the past decade: A personal view. *Linguistic Typology* 11 (1): 79–92.
- Crowley, Terry
 2007 *Field Linguistics: A Beginner’s Guide*. Oxford: Oxford University Press.
- Dalrymple, Mary, and Irina Nikolaeva
 2011 *Objects and Information Structure*. Cambridge: Cambridge University Press.
- Dixon, Robert M.W.
 1972 *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.
- Duranti, Alessandro
 1994 *From Grammar to Politics: Linguistic Anthropology in a Western Samoan Village*. Berkeley: University of California Press.

- Epps, Patience
 2008 *A Grammar of Hup*. Berlin / New York: Mouton de Gruyter.
 2009 Where differential object marking and split plurality intersect: Evidence from Hup. In *New Challenges in Typology. Transcending the Borders and Refining the Distinctions*, Patience Epps and Alexandre Arkhipov (eds.), 85–104. Berlin: Walter de Gruyter.
 2011 Linguistic typology and language documentation. In *The Oxford Handbook of Linguistic Typology*, Jae Jung Song (ed.), 634–649. Oxford: Oxford University Press.
- Evans, Nicholas
 2011 Semantic typology. In *The Oxford Handbook of Linguistic Typology*, Jae Jung Song (ed.), 504–533. Oxford: Oxford University Press.
- Gippert, Jost, Nikolaus Himmelmann, and Ulrike Mosel (eds.)
 2006 *Essentials of Language Documentation*. Berlin: Mouton de Gruyter.
- Greenberg, Joseph (ed.)
 1963 *Universals of Language*. Cambridge, MA: MIT Press.
- Haspelmath, Martin
 2007 Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11 (1): 119–132.
 2009 Pourquoi la typologie des langues est-elle possible? *Bulletin de la Société de Linguistique de Paris* CIV: 17–38.
 2010 Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86 (3): 663–687.
- Haspelmath, Martin, and Uri Tadmor (eds.)
 2009 *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: De Gruyter Mouton.
- Himmelmann, Nikolaus
 2006 Language documentation: What it is and what is it good for? In *Essentials of language documentation*, Jost Gippert, Nikolaus Himmelmann, and Ulrike Mosel (eds.), 1–30. Berlin: Mouton de Gruyter.
- Hjelmslev, Louis
 1970 *Language*. Madison, WI: University of Wisconsin Press.
- Hudson, Richard
 2007 Towards a useful theory of language. In *Proceedings of the Conference on Language Documentation and Linguistic Theory*, Peter Austin, Oliver Bond, and David Nathan (eds.), 3–11. London: School of Oriental and African Studies.
- Iemmolo, Giorgio
 2010 Topicality and differential object marking: evidence from Romance and beyond. *Studies in Language* 34 (2): 239–272.
- Joseph, Umbavu V.
 2007 *Rabha*. Leiden / Boston: Brill.

- Koch, Peter
 2001 Lexical typology from a cognitive and linguistic point of view. In *Language Typology and Language Universals. An International Handbook, Volume 2*, Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), 1142–1178. Berlin: Walter de Gruyter.
- Koptjevskaja-Tamm, Maria, Martine Vanhove, and Peter Koch
 2007 Typological approaches to lexical semantics. *Linguistic Typology* 11 (1): 159–185.
- Lazard, Gilbert
 2005 What are we typologists doing? In *Linguistic diversity and language theories*, Zygmunt Frajzyngier, Adam Hodges, and David Rood (eds.), 1–23. Amsterdam: John Benjamins.
 2006 *La quête des invariants interlangues: la linguistique est-elle une science?* Paris: Champion.
- Munro, Pamela
 2001 Field linguistics. In *The Handbook of Linguistics*, Mark Aronoff and Janie Rees-Miller (eds.), 130–149. Oxford: Blackwell.
- Newman, Paul, and Martha Ratliff (eds.)
 2001 *Linguistic Fieldwork*. Cambridge: Cambridge University Press.
- Nichols, Johanna
 2007 What, if anything, is typology? *Linguistic Typology* 11 (1): 231–238.
- Sapir, Edward
 1921 *Language. An Introduction to the Study of Speech*. New York: Harcourt & Brace.
- Schmid, Stephan
 this vol. Phonological typology, rhythm types and the phonetics-phonology interface. A methodological overview and three case studies on Italo-Romance dialects.
- Seiler, Hansjakob
 2001 The Cologne UNITYP Project. In *Language Typology and Language Universals, An International Handbook, Vol. 1*, Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), 323–344. Berlin: Mouton de Gruyter.
- Shopen, Timothy (ed.)
 2007 *Language Typology and Syntactic Description, 3 vols.* Cambridge: Cambridge University Press.
- Valenzuela, Pilar
 2011 Argument encoding and pragmatic marking of the transitive subject in Shiwilu (Kawapanan). *International Journal of American Linguistics* 77 (1): 91–120.

Wälchli, Bernhard

- 2007 Potential challenges of typologies based on primary data, illustrated with a case study on nominal plural. Paper read at the Pre-ALT Workshop on Linguistic Typology and Language Documentation. Paris, September 24-25.

Witzlack-Makarevich, Alena, Taras Zakharko, Lennart Bierkandt, and Balthasar Bickel

- 2010 Dissecting hierarchical agreement. Paper read at the workshop 'Referential hierarchy effects on the morphosyntax of verbal arguments,' Leipzig, September 28-29.

Zúñiga, Fernando

- 2007 The discourse-syntax interface in northwestern Amazonia. Differential object marking in Makú and some Tucanoan languages. In *Language Endangerment and Endangered Languages: Linguistic and Anthropological Studies with Special Emphasis on the Languages and Cultures of the Andean-Amazonian Border Area*, Leo Wetzels (ed.), 209-227. Leiden: Publications of the Research School of Asian, African, and Amerindian Studies (CNWS), University of Leiden.

- 2010 La marca diferencial del objeto en mapudungún. *Lingüística* 24: 141-164.

