Linguistic typology and language documentation

Patience Epps

University of Texas at Austin

0.1. Introduction

Linguistic typology and language documentation are closely aligned, even symbiotic endeavors. In its search to define the limits, patterns, and explanations that characterize cross-linguistic variation, typology's commitment to an empirical approach gives it a crucial dependence on the linguistic enterprise that provides these data. This is language documentation, the process of creating records of individual languages – that is, gathering a sample of data that represents a given language as fully as possible, and making this accessible to others, most notably non-speakers. In turn, the activities of documentation – determining the type of data to gather and the methods of doing so, and processing and representing these data for others – involve the application of cross-linguistically relevant constructs. Accordingly, they rely on typology.

In the past decade, documentary linguistics has emerged in its own right as a field of linguistic inquiry. It prioritizes the production of a rich, diverse, and extensive corpus, consisting primarily of naturally occurring discourse, and includes among its foci concerns of data preservation, accountability, interdisciplinary cooperation among researchers, and direct involvement of speech communities (Himmelmann 2006: 15). The goals of documentation are thus not identical to those of traditional linguistic description, which prioritizes the representation of the patterns underlying the data, and focuses on the production of grammars and dictionaries (Himmelmann 1998, Lehmann 2001, Woodbury

2003). Nevertheless, documentation and description are in practice inseparable, in that appropriate data collection and representation necessarily involve descriptive analysis, and vice versa.[1] In this paper, therefore, 'documentation' is understood as the broader endeavor that includes both data collection and analysis, while the term 'description' is reserved for the more analytical parts of the documentary process.

This chapter explores the close partnership between linguistic typology and language documentation. The discussion focuses first on the contributions of documentation to typology, particularly in shaping efforts to define universals and to make sense of linguistic diversity. Turning to the other side of the coin, it examines the importance of typology to the documentary enterprise, and explores typology's role in informing the representation of the language, ensuring its accessibility, and identifying ways in which it may be further refined. The final section considers future directions. Given the mutually constitutive nature of the relationship between these two areas of linguistic investigation, we may expect these to be fruitful.

0.2. The importance of documentation to typology

0.2.1. Issues of universality

The field of linguistic typology has traditionally focused on defining universals, the constraints on what is possible in human language. This endeavor is crucially empirical; typology bases its generalizations on broad cross-linguistic comparison. This, of course, is a principal point of contrast between typological and 'formalist' approaches to

language, associated with Chomsky's concept of universal grammar, which postulate universals on the basis of one or only a few languages (e.g. Chomsky 1980).

In typology, cross-linguistic generalizations may be phrased as absolute universals – i.e. exceptionless statements that hold true for all languages – but may also be conceived as tendencies (statistical universals), or as conditioned statements of the form 'if a language has X, it will also have Y' (implicational universals; Greenberg 1963). An emphasis on tendencies rather than absolutes has continued to gain priority in typological work – again in contrast to formalist approaches, which have focused on exceptionless universals in the attempt to define an innate language endowment (i.e., that which all children are assumed to bring to the learning of language).

Documentary linguistics is largely responsible for this shift of focus from absolute universals to tendencies. Ever since Greenberg (1986: 15) himself noted "the meagreness and relative triteness of statements that were simply true of all languages", the wealth of newly described languages entering the typological database has falsified claim after claim of universality (Bickel 2007: 242; see Plank and Filimonova 2000, Plank and Mayer 2006, Plank 2003). As Haspelmath (2007: 122) has observed, "almost every newly described language presents us with some 'crazy' new category that hardly fits existing taxonomies". True absolute universals of all kinds – even implicational ones – have proven to be exceedingly rare (Evans and Levinson, forthcoming).

There is no shortage of examples that illustrate the role of data from particular languages in stretching the limits of what was thought possible. A well-known case is the 'discovery' of object-initial word order in a number of Amazonian languages (Derbyshire 1977; see also, e.g., Munro 2003: 147, Olawsky 2007), which required a

substantial revision of our understanding of cross-linguistic word order constraints. Hajek (2007) calls attention to the unusual word-initial consonant clusters in the Austronesian languages Taba and Leti, which contradict several of Greenberg's (1978) generalizations about possible phonotactic structures. Blevins (2007) observes that the extensive discussion of sibilant harmony in the phonological literature – i.e. the phenomenon by which coronal sibilants within the word (or other domain) all agree with each other in terms of secondary place features (e.g. retroflex vs. non-retroflex) – rests mostly on descriptive data from a handful of North American languages. And Evans (2007: 24-25) notes the theoretical and typological significance of the remarkable system of multiple case-marking in the Australian language Kayardild, where case markers can stack on a single noun phrase, and can even contribute information relating to tense and mood in addition to syntactic role.

Criticisms that attention to such rara amounts to no more than 'butterfly collecting' (Chomsky 1998 [1977/1979], cf. Johnson 2002) are thus clearly unfounded. It is the documentation of individual languages, and the attention to specific features of these languages, that enable typologists – and other linguists – to shape our definition of human language, as well as to predict linguistic patterns that are not yet attested (Blevins 2007: 10). Moreover, many of these data come from endangered languages, a fact which underscores the importance of recording these languages as quickly and as thoroughly as possible (and, where possible, helping to revitalize them). The attention to language endangerment is a critical component of the documentary initiative; it is likewise important to typology, since only living languages will ensure a continued source of data for cross-linguistic comparison.

In addition to contributing data to test cross-linguistic generalizations, language documentation is a prerequisite for a balanced representation of the world's languages, one that takes into account the full range of genealogical (genetic) and geographic diversity. Typologists are well aware that sampling is subject to areal bias, and that limiting this requires that each of the world's major geographical regions be well represented in the sample (Dryer 1989, 1992). Yet our current knowledge of the world's languages is far from equally distributed. Whole areas of the world are represented by languages that are highly exotic from the perspective of most linguists, in comparison to the more thoroughly studied languages of Europe. Amazonia and New Guinea represent two of the most extreme cases: in these regions, linguistic diversity is remarkably high (both in terms of numbers of languages and their relation to each other),[2] but documentation is minimal – and many of the languages are highly endangered.

Nevertheless, the past two decades have seen a remarkable burst of documentary studies in many of these chronically under-represented regions. The impact on typology has been considerable, leading to substantial revisions of earlier generalizations. Dixon and Aikhenvald (1999:1) note that Amazonian languages have been a particularly rich source of typological surprises: "In case after case, just as [Dixon] thought he had achieved some significant typological statement, a counter-example popped up; and this was invariably from a language of Amazonia". The example of object-initial word order, cited above, is a case in point; if linguists had been as familiar with Amazonian languages as they were with European languages at the time of Greenberg's (1963) ground-breaking survey, it is likely that object-initial word order would never have been thought to be so unusual, let alone impossible. A similar example concerns Greenberg's generalizations

about subject-verb order inversion in questions; these were based entirely on languages of Europe (nearly one-fourth of his sample), but in fact, subject-verb order inversion is extremely rare outside this region. Of the few non-European languages that do exhibit this typologically unusual strategy, at least two Amazonian languages – Hup (Nadahup/Makú family, Epps 2008) and Warekena (Arawak family, Aikhenvald 1998) – violate Greenberg's proposed implicational universal that inversion of subject and verb occurs in yes-no questions only if it also occurs in interrogative word questions.

As this discussion has illustrated, linguistic typology recognizes that our generalizations about human language are crucially dependent on data from a wide range of individual languages. Nettle and Romaine (2000: 10-11, see Austin and Simpson 2007: 6) observe that "to exclude exotic languages from our study is like expecting botanists to study only florist shop roses and greenhouse tomatoes and then tell us what the plant world is like". Language documentation is the essential means by which these languages may be included.

0.2.2. Issues of diversity

As documentary data from more and more languages become available, typologists are gradually coming to grips with the remarkable extent to which languages may vary. As Evans and Levinson (forthcoming) argue, the extraordinary extent of this diversity has profound implications: the human capacity for language may well be far more complex than the innate, invariant, and finite faculty proposed by Chomsky. At the same time, an expanded data set enables us to revise and refine our understanding of the patterns and tendencies that do exist. Rather than characterize diversity according to "sharp

boundaries between possible and impossible languages, between sharply parameterized variables, or by selection from a finite set of types", diversity is better represented as "clusters around alternative architectural solutions, by prototypes (like 'subject') with unexpected outliers, and by family-resemblance relations between structures ('words', 'noun phrases') and inventories ('adjectives')" (Evans and Levinson, forthcoming: 43). In other words, as Bickel (2007: 245) puts it, "linguistic diversity is captured by large sets of fine-grained variables, not by grand type notions".

The emerging awareness that a more sophisticated understanding of language requires a closer attention to variation, detail, and complexity calls for an ever more fine-grained approach to typology (see Bickel 2007: 247). Such an approach is exemplified in a number of recent contributions to the field, many of which ground their discussion in a corpus of primary data from a newly documented language. For example, Seifart (2009) observes that the system of nominal classification in Miraña, a Bora-Witotoan language of Amazonia, defies characterization according to the limited set of prototypes established for classifier systems cross-linguistically. Accordingly, he argues for the need to "shift the focus away from broad types defined on the basis of few characteristics as the basis for typological characterization and comparison towards a model with uses a larger number of more detailed and more varied parameters" (Seifart 2009: 365). Similarly, Schultze-Berndt (2007) discusses how a detailed investigation of the expression of motion events in Jaminjung (a non-Pama-Nyungan language of Australia) informs a revision of Talmy's (1985, 1991) typology. While Talmy distinguishes between 'satellite-framed' languages (in which path is expressed lexically as a 'satellite' to the motion verb) and 'verb-framed' languages (in which both path and motion are

lexicalized in the verb itself), the Jaminjung data suggest that concepts like 'satellite', 'path', and 'manner' require more nuanced definitions (see also O'Connor 2009 for Lowland Chontal of Oaxaca, Mexico).

Discussions like those of Seifart and Schultze-Berndt highlight the mutually constitutive nature of documentation and typology. A more fine-grained typological understanding necessarily rests on a comparably fine-grained catalogue of cross-linguistic variation, which in turn requires more detailed descriptions of individual languages, informed by knowledge of the relevant parameters on the part of the linguists doing the documentation. As Khanina (2009) illustrates with the case of desideratives – expressions of wanting – a refined typology crucially informs the quality of description and documentation, and vice versa.

The goal of refining our typological focus is well served by contemporary documentary methodology, which stresses the collection of a large and diverse corpus. The corpus is expected to capture an extensive range of data types, spanning genres, styles, registers, speakers, etc., and presenting these in multimedia formats. In addition to detailed descriptive records, it includes ethnographic and sociolinguistic notes (see Himmelmann and Evans 2007, Lehmann 2001, Franchetto 2006). The principal focus of documentation is naturally occurring discourse (see Woodbury 2003, Sherzer 1990), but many linguists agree that elicitation is also a valuable source of data (Chelliah 2001), as are 'hybrid' techniques that bridge naturalistic data collection and structured elicitation (Bowerman 2007).

The extensive diversity and depth of the documentary corpus increases the likelihood that it will be a relatively complete representation of the spoken language, and

one that can inform fine-grained typological investigation. The study of evidentiality, the grammaticalized expression of information source, provides an excellent example of this point. Speakers often omit evidentials altogether in elicited utterances (Aikhenvald 2004: 18); as Seki (2000: 347, in Aikhenvald 2004: 18) describes for Kamaiurá (Tupi-Guarani), this may result in sentences that native speakers find "artificial, sterile, deprived of colour". On the other hand, Lidz (2007) notes that a full investigation of a complex evidential system may only be possible with some elicitation, as in the case of Yongning Na (Tibeto-Burman), in which evidentiality interacts with verbal semantics and a conjunct/disjunct-like system. Even in natural discourse, the use of evidentials may be highly variable; for example, in Hup (Nadahup), nonvisual and inferred evidentials are common in spontaneous conversation, but are quite rare in narrative (which uses primarily the reported specification; Epps 2008: 641-654). Similarly, Michael (2008) shows that speakers of Nanti (Arawak) use more evidentials when concerned about personal responsibility for an event, and fewer evidentials in other situations.

While typological work to date has drawn primarily from descriptive grammars, the extensive corpus generated through documentation – and made available by archiving – promises to be a far richer source of data for typology. The documentary corpus makes it possible for typologists to compensate for gaps in grammars, and allows for doublechecking and second opinions (Wälchli 2007, Blevins 2007: 4). Moreover, as we saw for evidentials, a diverse corpus provides information on language-internal variation – not only across genres, styles, speakers, and discourse contexts, but also across linguistic structures. This last point may be of particular interest to typologists, in light of Bresnan's (2007) recent observation that many patterns of variation in language-internal

preferences for particular structures show significant correlations with similar patterns of cross-linguistic variation. For example, Bresnan (2007: 300) observes that variation in the choice of the double object construction vs. an indirect object for certain verbs in English (e.g. '*Ted denied Kim the opportunity to march'* vs. '*Ted gave Joey permission to march, but he denied it to Kim'*), depending on whether the sequences of objects involve nouns or pronouns, mirrors a very similar cross-linguistic pattern described by Haspelmath (2004).

Such information on structural variation is often missing from grammars, which tend to present a normalized view that neglects unsystematic or infrequent structures (Wälchli 2007); likewise, marginal structures may also be overlooked if our techniques of data collection are too restricted (Bresnan 2007). Accordingly, Bresnan (2007: 302) stresses "the need to support claimed generalizations with multiple empirical sources of converging evidence, including observations of ecologically natural language use". This need holds true equally for typology and for linguistic description. Both approaches to linguistic analysis seek to generalize across a particular domain, as Moravcsik (2007: 34) observes – typology across languages, description across utterances within a single language. Just as typological investigation is moving toward a more fine-grained, detailed approach to cross-linguistic variation, so will the move toward a richer and more thorough documentation result in finer-grained analyses of individual languages – thereby enriching the continued partnership between the two enterprises.

In addition to fostering a more nuanced approach to typological questions, the growing attention to linguistic diversity is shaping a new focus in the discipline: Rather than simply attempting to define what is universal in human language, typologists are

now turning to explanations for why diversity is the way it is – "what's where why?" (Bickel 2007: 239). This question demands the consideration of multiple variables: not only universal preferences, but also geographic and genetic (genealogical) distributions, diachronic change, and the interaction between language and social, cognitive, and cultural factors. This last consideration, in particular, entails a clean break with assumptions that have shaped much of linguistic thinking in the past decades, that KNOWLEDGE of language (Saussure's 'langue', Chomsky's 'competence', limited primarily to grammar and lexicon) can and should be studied independently of its USE (Saussure's 'parole', Chomsky's 'performance'). Rather, it posits that "observed structures arise, through time, by summing the outcomes of many communicative acts by individuals" (Evans 2003: 15; see also Keller 1994, Haspelmath 1999).

Again, the priorities of documentary linguistics closely parallel this emerging concern of typologists. Contemporary documentary methodology emphasizes that data 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; see also Lehmann 2001: 90, Hill 2006). Such an organic approach is essential if we are to achieve a more complete typological understanding of diversity. While the role of cultural context in shaping linguistic structures has long been overlooked, it is illustrated in a number of recent discussions of typologically noteworthy phenomena in particular languages. For example, Evans (2003) observes that kin-sensitive pronouns and dyadic kin terms in some Australian languages (e.g. 'they two' [= husband and child of speaker], 'father and child') illustrate the intrinsic link between these linguistic structures and their cultural context. Similarly, in

the Amazonian language Hup (Nadahup), most nouns referring to generic types of human beings ('child', 'shaman', 'youth', 'woman') must be preceded by another noun (typically either an ethnic denomination or a default third person singular pronominal form). Elsewhere in the language, this 'bound' construction is associated with inalienable possession (e.g. kin terms) or the inherent relationship between a part and a whole (e.g. plant parts); its occurrence with generic human nouns is typologically unusual, but may derive from the cultural emphasis on the person as inherently associated with a social group (Epps 2008: 258-259).

This holistic view of language as inseparable from its social and cultural context has implications not only for explanations of WHY the objects of typological investigation exist, but also for determining WHAT should be investigated in the first place. As Rumsey (2007) points out, the focus of typology has traditionally been 'langue' (i.e. lexicon and grammar), but considerations of 'parole' may be equally interesting areas of study. For example, parallelism, or structured repetition in discourse, is an important component of verbal artistry in many of the world's languages; moreover, its basis in notions of 'equivalence' can inform our understanding of grammar and semantics, both language-internally and cross-linguistically (Rumsey 2007; see also Fox 1977, Jakobson 1960). Similarly, Evans (2007) considers the typological and sociolinguistic relevance of the special registers (initiation, respect, etc.) encountered in Australian languages, and notes the semantic insights they bring to the study of the corresponding 'everyday' forms (Evans 2007: 34). Even for those linguistic phenomena that have traditionally been considered elements of 'langue', a fine-grained typological understanding can hardly ignore considerations of language in use. This is certainly the case for evidentials, for

which realization and function depend heavily on usage, as discussed above. It also applies to personal pronoun systems, which relate to categories of possible kinds of speaker and addressee, and the relationships among them (Rumsey 2007, Evans 2003); and likewise to lexical phenomena such as idioms (Thurston 2007). It 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 (Rumsey 2007, Evans 2007: 34).

0.3. The importance of linguistic typology to documentation

0.3.1. Issues of representation

Making a language accessible to non-speakers is a crucial commitment of documentation – indeed, this is its primary purpose (Lehmann 2001: 86). This activity requires representation, the elaboration of what Himmelmann and Evans (2007) call 'raw data' – recordings of communicative events – to produce 'primary data' – transcriptions, translations, and linguistic and ethnographic commentary. Representation rests on analysis (see Lehmann 2001, Woodbury 2003), the exposition and explanation of patterns in the data – i.e. the descriptive part of the endeavor – which in turn shapes the ongoing development of the corpus itself (Woodbury 2003: 42; see §0.1 above). Finally, analysis, by definition, entails an abstraction from the language itself: this involves categorization, which makes it possible to represent the infinite number of linguistic utterances in finite and relatively concise terms, and it involves a metalanguage, which allows us to effectively communicate these generalizations to others.

In other words, efficient representation requires typology. Particularly in communicating patterns, but also in discovering them, we rely on constructs that have some degree of relevance beyond the language itself. For example, accurate transcription requires phonemic representation: to represent a particular phoneme as /t/, we have to define its instantiation as a phonetic category in the language in question, and recognize that this category shares certain (though probably not all) features with those represented as /t/ in a variety of other languages.

Yet there is considerable flexibility in the extent to which representation must rely on cross-linguistically defined concepts, as evidenced by early structuralist experiments in describing languages entirely 'in their own terms' (see Boas 1911, Dryer 1997, Haspelmath 2007). These efforts led, for example, to Garvin's (1948) classification of Kutenai word stems into the classes 'W', 'X', and 'Y' on morphological grounds – even though W and X correspond semantically to what are traditionally called 'nouns' and 'verbs' (see Dryer 1997: 117).

Most linguists agree that Garvin's solution is not ideal. It is unwieldy, in that it requires us to learn a new metalanguage for each language under investigation; it inhibits forming and testing predictions about the internal interactions among subsystems of the languages; and it obscures not only the similarities between this language and others, but also the potentially interesting points of variation. Yet typologists are also well aware that, on some level, Garvin's solution is reasonable in that it recognizes that structures such as word classes are not identical across languages (Dryer 1997: 117). This is a basic problem for typology: we need to be able to compare structures across languages, but these structures vary from language to language (e.g. Croft 1991, 2000; Dryer 1997;

Haspelmath 2007, 2008; Stassen, this volume). Likewise, it is a problem for linguistic description: we cannot have confidence in our metalanguage – we cannot reasonably call something a 'verb' or an 'adjective' – without accepting that these categories have some cross-linguistic validity.

Typology offers a solution to this problem. While particular categories like 'verb', 'adjective', and 'subject' are not invariant or even necessarily attested across all languages, typologists have recognized that they are useful descriptive labels that capture real tendencies. In other words, they represent cross-linguistically relevant categorial prototypes (Croft 1991, 2000, 2007; Dryer 1997; Haspelmath 2007; Evans and Levinson forthcoming). This has profound implications for linguistic description: Just as the documentation of particular languages has allowed us to generalize more accurately across languages, these generalizations in turn give us the tools to effectively represent the structures found in particular languages.

While an understanding of cross-linguistic similarities is important for coherent description, an awareness of how languages differ is likewise essential. It is by now a commonplace in the descriptive and documentary literature that our native and familiar languages can act as blinders, preventing us from understanding a language on its own terms (e.g. Gil 2001, Blevins 2007). Examples of such skewed representations abound; for example, many grammatical descriptions written prior to the twentieth century present languages having no morphological case (such as English) according to the six-case Latin model (ablative, genitive, vocative, etc.; see Haspelmath 2008, Gil 2001). Similarly, French priests describing the Nadahup languages of Amazonia, in which nouns are not usually marked for gender, exaggerated the frequency of masculine and feminine class

terms with animate nouns, in explicit parallel to the grammatical gender of Romance languages (Rivet et al. 1925). An awareness of just how diverse languages can be – a perspective contributed by typology – helps us to understand familiar patterns as only one option among many. It encourages us to be open to the new patterns we encounter in the documentary process, to meet Hockett's (1993: 4, in Blevins 2007: 4) challenge to "let [the language] show us how it works – instead of trying to force matters into some conceptual frame of reference we have imported, perhaps without realizing it, from elsewhere" (see also Gil 2001).

Most contemporary descriptive work has clearly progressed far beyond the use of Latin or French as an explicit model for representing an undocumented language. Nevertheless, the lessons of the past also remind us that – just as the documentation of the world's existing languages is far from complete – the typological "deconstruction of the Eurocentric metalanguage" (Daniel 2007: 74) is an ongoing process. Thus while typology gives us valuable tools for understanding and representing a particular language, these tools should not be applied uncritically. Approaches that rely too heavily on a predetermined typological 'template' risk overlooking or obscuring interesting features of the language, as noted by Baerman and Corbett (2007) for Tübatulabal (Uto-Aztecan) aspect marking, and by Lüpke (2007: 187) for Jalonke (Mande) voice. Gil (2001: 126-128) observes the tradeoff between a "bottom-up" approach, which starts with the data and allows categories and structures to emerge, and a "top-down" approach, which relies on a predetermined set of grammatical categories. Any descriptive or documentary effort will necessarily involve both of these approaches to some degree, and it is up to the fieldworker to strike the appropriate balance.

0.3.2. Documentation, typology, and 'theory'

The tension between the general and the particular – their role both in the representation of individual languages and in our understanding of what constitutes human language more generally – underscores the inseparable relationship between theory and the representation of a particular language. This is particularly clear in the case of descriptive analysis; as Gil (2001: 126) observes, description requires proposing categories and formulating generalizations about them, just as any empirically grounded theory must account for a range of facts (see also Dixon 1997). However, this is also true of the documentation process more broadly; even the collection of raw data must be theoretically informed if it is to result in a relatively complete corpus, representative of the various grammatical structures, genres, styles, and other linguistic resources available to speakers.

Nevertheless, a widespread perception persists among linguists that 'theory' and 'description' are "fundamentally separate" (Eastman 1978: 3), even diametrically opposed. As Van Valin (2007: 253) observes, the history of this opposition in the linguistics of the twentieth century has been a strained one, marked by the "conflict between the desire to capture the 'structural genius' of languages (Sapir 1921) and the desire to capture what is universal in human language" (see also Gil 2001: 126). However, much of this can be attributed to an assumption that 'theory' in linguistics is synonymous with 'formalism', i.e. a type of approach that involves a highly specific metalanguage, with the goal of producing an abstract model of speaker knowledge. Yet 'theory' is better understood as a mechanism for reducing complexity by means of

generalizations (e.g. Gil 2001: 126), or as a set of explanations for why languages (or their particular features) take the form they do (Dryer 2006a), and is thus by no means the exclusive domain of formalists (see also Dixon 2009, Blevins 2007: 4).

The theoretical framework employed in most documentary and descriptive work is essentially the same approach that is taken by most typologists. This framework, which is grounded largely in traditional grammar and is not associated with any particular formalism, has been given the label 'basic linguistic theory' (Dixon 2009; Dryer 2001, 2006a, 2006b). Its development depends on formulating generalizations that are empirically grounded in the properties of particular languages, but that go beyond these to capture broader cross-linguistic facts. This, of course, is the domain of typology, informed by documentation – and typology has had the most influence on the development of basic linguistic theory in recent decades (Dryer 2001, Nichols 2007: 235).

0.3.3. Refining and enriching documentary linguistics

As this discussion has made clear, linguistic typology and documentary linguistics share much of the same architecture: a common theoretical framework, an awareness of cross-linguistic similarity and variability, and a goal of forming and representing generalizations over diverse realizations (whether cross-linguistically or language-internally; see Moravcsik 2007). This common ground ensures that, just as documentation informs typology, typology profoundly informs documentation, probably more so than any other linguistic subfield. Typology contributes to the development of a precise metalanguage for the representation of particular languages; it suggests

predictions to be tested within a given language, such as interactions among logically independent parameters like word order; and it helps draw our attention to gaps in a language's descriptive representation and documentary corpus. Furthermore, when a particular phenomenon does not fit a cross-linguistically established category, typology helps us to see what is noteworthy and of possible theoretical interest, and what should be explored through further documentation.

The fertile interaction between these disciplines is frequently noted. Dryer, for example, observes that "looking at lots of languages gives one a sense of what languages are like", so much so that a typologist is likely to have insights into a particular language that even a specialist in the language (who is not typologically trained) will not have (Dryer 2002: 18; see Song 2007: 16). Nichols (2007: 235) stresses the usefulness of "the framework-neutral definitions, the growing body of substantive knowledge, and statistical and probabilistic knowledge [of typology], all of which are readily applicable to description, comparison, and pinpointing what is distinctive and valuable about a particular language".

A particularly valuable contribution of typology to documentation is the recognition that language-specific categories are variable, and may therefore only roughly match cross-linguistically defined prototypes and inventories (see §0.3.1 above). This awareness encourages more sophisticated and detailed description; as Haspelmath (2007: 128) observes, "by shedding the assumption of *a priori* categories, descriptive linguists can avoid getting into category-assignment controversies and can concentrate on refining their descriptions." In other words, rather than simply calling something an 'adjective' and moving on – or worrying at length about whether it should be called an

adjective – the linguist will be led to consider in detail the morphosyntactic, semantic, and other properties that define this category within the language in question, how they resemble those of adjectives in other languages, and how they differ.[3] This awareness also helps documentary linguists avoid being stymied by a lack of confidence, a concern that their analysis may be flawed because it does not fit preconceived notions.

The typological approach also gives us the tools to understand and explain WHY a language-particular category may be non-prototypical. In addition to considering extra-linguistic elements of explanation (i.e. rooted in the culture, history, and environment of the speakers, see §0.2.2 above), typology takes a 'dynamic' perspective – a view that linguistic systems are always undergoing change, that one prototypical language state may over time develop into another, and that particular instantiations of a grammatical feature in a given language may represent any point in this historical transition (e.g. Croft 2003a, 2003b). This perspective helps us to make sense of phenomena that appear to fall somewhere in between two cross-linguistically defined prototypes. For example, a number of Amazonian languages, such as the sisters Hup and Yuhup (Nadahup; Epps 2008, Ospina 2002), and Apurinã (Arawak, Facundes 2000), have systems of 'bound' nouns that serve a classifying function, but are neither pervasive enough in the lexicon nor morphologically distinct enough from normal nouns to be easily defined as nominal classifiers. However, these nominal forms can be understood as representing a language state somewhere between a prototypically non-classifier system and a full-fledged, grammaticalized set of classifiers (see Dixon 1986; Grinevald 2000, 2002; Grinevald and Seifart 2004). In Hup, moreover, the 'incipient' nature of the system is particularly clear in that the set of 'class terms' occurs primarily with a restricted set of lexical items,

neologisms referring to items of non-native manufacture (Epps 2007). It is the dynamic approach of typology that allows us to make sense of this system; the result is a description that is accessible to others, is cross-linguistically relevant yet richly language-specific, and that allows the prediction of likely avenues of change in the future.

Just as categories defined within a particular language will vary with respect to cross-linguistically defined prototypes, a typological approach also attends to language-internal variation in category membership. This gives documentary linguists the flexibility to recognize and make sense of entities within the language that do not fit the language-internal prototype any better than they fit the cross-linguistic one. For example, individual morphemes undergoing grammaticalization may display characteristics associated with two different lexical or grammatical categories within the language in question (e.g. evidentials developing from verbs; see Aikhenvald 2004: 271-275). In other cases, this sort of liminality is semantically motivated and therefore relatively stable; for example, words referring to periods of time often have features of both nouns and verbs (e.g. Hup *wag* 'day' and *j'əb* 'night' normally appear as arguments of verbs and can take most nominal morphology, but also can occur as predicates and take verbal inflection; Epps 2008: 163). Once again, attention to such details results in a more refined description of the language being documented (Haspelmath 2007, 2008).

Finally, in addition to contributing to a theoretically sophisticated model for linguistic description and documentation, typology also informs the relatively framework-neutral terms of its presentation. This is critical to its accessibility. Whereas descriptive work couched in a particular formalism (e.g. generative grammar, tagmemics, etc.) tends to be accessible to relatively few people and to quickly become dated,

typologically informed work is more likely to remain useful not only to generations of linguists, but to speakers and their descendents as well.

0.4. A continuing partnership

The past decade has seen linguistic typology come into its own as a mature discipline (see, for example, Bickel 2007 and associated papers in *Linguistic Typology* 11). Likewise, the past decade has seen a renewed interest in the description of little-known and endangered languages, and the emergence of documentation as a focus of investigation in its own right (Himmelmann 1998, Lehmann 2001, Woodbury 2003, Gippert et al. 2006, Austin et al. 2007, etc.). The parallel expansion of these two subfields is not a coincidence: The emerging awareness of the theoretical importance of linguistic diversity has fueled interest in typology, and has at the same time made documentation a priority of the field – especially as the problem of language endangerment has gained widespread attention (Hale et al. 1992; see also Nichols 2007: 235, Woodbury 2003: 37). In the process, typology and documentary linguistics have contributed greatly to each other. Documentation has shaped new paths of typological investigation, ranging from what is to be investigated to how patterns are to be explained. Typology has given documentary linguists a solid and accessible theoretical structure on which to build their discussions of individual languages.

As the partnership between these two subfields of linguistics has grown closer and more productive, it has also brought into focus their mutual dependence. For typology, continued progress depends on access to data from as many languages as possible, spanning as many modalities, registers, styles, etc. as possible. For any given

documentary effort to be of maximum usefulness to typologists, the corpus must be large, rich, and diverse, built on multiple methods of data collection. It must be ethnographically grounded, such that the record involves 'thick' rather than 'thin' description (Plank 2007: 47, cf. Geertz 1973; see also Evans and Levinson forthcoming, Rumsey 2007, Evans 2007). In addition to completeness, documentary integrity is also critical: an accurate understanding of the limits of linguistic diversity requires that the descriptions of phenomena in individual languages be neither exaggerated nor understated (e.g. Lehmann 2001: 92, Song 2007: 39-40).

For language documentation, in turn, the goal of more refined, detailed, and theoretically sophisticated records of individual languages will require comparable progress in refining our typological understanding of particular linguistic phenomena, and in accurately characterizing the range of possible structures that may be encountered in any given case. Furthermore, these discussions need to be made available and accessible to documentary linguists. Materials such as Shopen (2007) that offer detailed and typologically informed discussions of a wide range of phenomena are invaluable resources. There is no doubt that documentary work would benefit from more such comprehensive, in-depth typological studies of language.

The association between linguistic typology and language documentation has developed into a rich and productive symbiosis, and its continued fruitfulness will be guaranteed by each subfield's increasing participation in the other. As documentary linguists gain more training and experience in typology, they will become more aware of the kinds of questions typologists are asking, of the theoretical and typological significance of the phenomena they are describing, and of the level of descriptive detail

necessary to answer them. Documentary linguists must confront the considerable challenge handed them by typologists: to produce descriptions in formats that will enable and facilitate comparison across languages (Plank 2007: 46), but also to remain true to the languages themselves, without forcing them into ill-fitting predetermined categories (Gil 2001; Haspelmath 2007, 2008). For their part, typologists should be aware that the documentary corpus may give considerable insight into the intricacies and complexities that characterize particular linguistic structures, but which may be oversimplified in a typological overview that draws from only a small part of that corpus (e.g. a page in a grammar sketch) – even so far as to compromise the final typological conclusions. Finally, as more and more languages become endangered or moribund, typologists must continue to recognize the importance and urgency of language documentation, and should join in on the documentary endeavor whenever possible.

Further reading

Ameka, F., Dench, A., and Evans, N. (eds.). (2006). *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter.

Austin, P., Bond, O., and Nathan, D. (eds.). (2007). *Proceedings of the Conference on Language Documentation and Linguistic Theory*. London: Department of Linguistics, SOAS. Online at http://www.hrelp.org/publications/ldlt/papers/ldltproceedings.html.

Austin, P. and Simpson, A. (eds.). (2007). *Endangered Languages*. Hamburg: Helmut Buske Verlag.

Epps, P. and Arkhipov, A. (eds.). (2009). *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*. Berlin: Mouton de Gruyter.

Gippert, J., Himmelmann, N., and Mosel, U. (eds.). (2006). *Essentials of Language Documentation*. Berlin: Mouton de Gruyter.

Himmelmann, N. (1998). 'Documentary and Descriptive Linguistics', *Linguistics* 36: 161-95.

Newman, P. and Ratliff, M. (eds.). (2001). *Linguistic Fieldwork*. Cambridge: Cambridge University Press.

Woodbury, A. (2003). 'Defining Documentary Linguistics', in P. Austin (ed.), *Language Description and Documentation*, vol. 1, 35-51. (Endangered Languages Project.) London: School of Oriental and African Studies.

References

Aikhenvald, A. (1998). 'Warekena', in D. Derbyshire and G. Pullum (eds.), *Handbook of Amazonian Languages*, vol. 4, pp. 225-439. Berlin: Mouton de Gruyter.

---- (2004). *Evidentiality*. Oxford: Oxford University Press.

Austin, P., Bond, O., and Nathan, D. (eds.). (2007). *Proceedings of the Conference on Language Documentation and Linguistic Theory*. London: Department of Linguistics, SOAS. Online at http://www.hrelp.org/publications/ldlt/papers/ldltproceedings.html.

---- and Simpson, A. (2007). 'Introduction', in P. Austin and A. Simpson (eds.), *Endangered Languages*, 5-18. Hamburg: Helmut Buske Verlag.

Baerman, M. and Corbett, G. (2007). 'Linguistic Typology: Morphology', *Linguistic Typology* 11: 115-117.

Bickel, B. (2007). 'Typology in the 21st Century: Major Current Developments', *Linguistic Typology* 11: 239-251.

Blevins, J. (2007). 'Endangered Sound Patterns: Three Perspectives on Theory and Description', *Language Documentation and Conservation* 1: 1-16. Online at http://nflrc.hawaii.edu/ldc.

Boas, F. (1911). 'Introduction', in *Handbook of American Indian Languages*, part I, 5-83. Washington: Bureau of American Ethnology.

Bowerman, M. (2007). 'Elicited Production: Tools for Semantic Typology', paper presented at the Pre-ALT workshop on Linguistic Typology and Language Documentation, 24-25 September, Paris, France.

Bresnan, J. (2007). 'A Few Lessons from Typology', *Linguistic Typology* 11: 297-306.

Bybee, J. (1985). *Morphology: A Study of the Relation between Meaning and Form.* Amsterdam: John Benjamins.

Chelliah, S. (2001). 'The Role of Text Collection and Elicitation in Linguistic Fieldwork', in P. Newman and M. Ratliff (eds.), *Linguistic Fieldwork*, 152-165. Cambridge: Cambridge University Press.

Chomsky, N. (1980). 'On Cognitive Structures and Their Development: A Reply to Piaget', in M. Piattelli-Palmarini (ed.), *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*, 35-52. Harvard University Press.

---- (1998). 'Language and Responsibility', in *On Language*, 1-197. New York: The New Press. (Original work published 1977/1979.)

Comrie, B. (1976). *Aspect.* Cambridge: Cambridge University Press.

Croft, W. (1991). *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: University of Chicago Press.

---- (2000). 'Parts of Speech as Language Universals and as Language-Particular Categories', in P. Vogel and B. Comrie (eds.), *Approaches to the Typology of Word Classes*, 65-102. Berlin: Mouton de Gruyter.

---- (2003a). 'Typology', in M. Aronoff and J. Rees-Miller (eds.), *The Handbook of Linguistics*, 337-368. Oxford: Blackwell Press.

---- (2003b). *Typology and Universals,* 2nd edition. Cambridge: Cambridge University Press.

---- (2007). 'Typology and Linguistic Theory in the Past Decade: A Personal View', *Linguistic Typology* 11: 79-91.

Daniel, M. (2007). 'Representative Sampling and Typological Explanation: A Phenomenological Lament,' *Linguistic Typology* 11: 69-78.

Derbyshire, D. (1977). 'Word Order Universals and the Existence of OVS Languages', *Linguistic Inquiry* 8: 590-599.

Dixon, R. M. W. (1986). 'Noun Classes and Noun Classification in Typological Perspective', in C. Craig (ed.), *Noun Classes and Categorization*, 106-112. Amsterdam: John Benjamins.

---- (1997). *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

---- (2009). *Basic Linguistic Theory.* Oxford: Oxford University Press.

---- and Aikhenvald, A. (1999). 'Introduction', in R. M. W. Dixon and A. Aikhenvald (eds.), *The Amazonian Languages*, 1-21. Cambridge: Cambridge University Press.

Dryer, M. (1989). 'Large Linguistic Areas and Language Sampling', *Studies in Language* 13: 257-292.

---- (1992). 'The Greenbergian Word Order Correlations', *Language* 68: 81-138.

---- (1997). 'Are Grammatical Relations Universal?' in J. Bybee, J. Haiman, and S. Thompson (eds.), *Essays on Language Function and Language Type*, 115-143. Amsterdam: John Benjamins.

---- (2001). 'What is Basic Linguistic Theory?' Online at http://linguistics.buffalo.edu/people/faculty/dryer/dryer/blt.

---- (2002). 'Minute Details, the Big Picture, and Joseph Greenberg,' *Linguistic Typology* 6: 16-18.

---- (2006a). 'Functionalism and the Theory-Metalanguage Confusion', in G. Wiebe, G. Libben, T. Priestly, R. Smyth, and S. Wang (eds.), *Phonology, Morphology, and the Empirical Imperative: Papers in Honour of Bruce Derwing*, 27-59. Taipei: The Crane Publishing Company.

---- (2006b). 'Descriptive Theories, Explanatory Theories, and Basic Linguistic Theory', in F. Ameka, A. Dench, and N. Evans (eds.), *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter.

Eastman, C. (1978). *Linguistic Theory and Language Description.* Philadelphia: J. B. Lippencott Co.

Epps, P. (2007). 'Birth of a Noun Classification System: The Case of Hup', in L. Wetzels (ed.), *Language Endangerment and Endangered Languages: Linguistic and Anthropological Studies with Special Emphasis on the Languages and Cultures of the Andean-Amazonian Border Area*, 107-128. (Indigenous Languages of Latin

America series [ILLA].) Leiden University, The Netherlands: The Research School of Asian, African, and Amerindian Studies (CNWS).

---- (2008). *A Grammar of Hup*. (Mouton Grammar Library 43.) Berlin: Mouton de Gruyter.

Evans, N. (2003). 'Context, Culture, and Structuration in the Languages of Australia', *Annual Review of Anthropology* 32: 13-40.

---- (2007). 'Warramurrungunji Undone: Australian Languages in the 51st Millenium,' in P. Austin and A. Simpson (eds.), *Endangered Languages*, 19-44. Hamburg: Helmut Buske Verlag.

---- and Levinson, S. (Forthcoming). 'The Myth of Language Universals: Language Diversity and its Importance for Cognitive Science', *Behavioral and Brain Sciences*.

Facundes, S. (2000). *The Language of the Apuriña People of Brazil*, Ph.D. diss., University of New York at Buffalo.

Fox, J. (1977). 'Roman Jakobson and the Comparative Study of Parallelism', in van C. H. Schooneveld and D. Armstrong (eds.), *Roman Jakobson: Echoes of his Scholarship*, 59-90. Lisse: Peter de Ridder Press.

Franchetto, B. (2006). 'Ethnography in Language Documentation', in J. Gippert, N. Himmelmann, and U. Mosel (eds.), *Essentials of Language Documentation*, 183-212. Berlin: Mouton de Gruyter.

Garvin, P. (1948). 'Kutenai III: Morpheme Distributions (Prefix, Theme, Suffix)', *International Journal of American Linguistics* 14: 171-187.

Geertz, C. (1973). 'Thick Description: Toward an Interpretive Theory of Culture', in *The Interpretation of Cultures: Selected Essays*, 3-30. New York: Basic Books.

Gil, D. (2001). 'Escaping Eurocentrism,' in P. Newman and M. Ratliff (eds.), *Linguistic Fieldwork*, 102-132. Cambridge: Cambridge University Press.

Gippert, J., Himmelmann, N., and Mosel, U. (eds.). (2006). *Essentials of Language Documentation*. Berlin: Mouton de Gruyter.

Greenberg, J. H. (1963). 'Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements', in J. H. Greenberg (ed.), *Universals of Language*, 73-113. Cambridge, MA: MIT Press.

---- (1978). 'Some Generalizations Concerning Initial and Final Consonant Clusters,' in J. Greenberg (ed.), *Universals of Human Language*, 243-279. Stanford: Stanford University Press.

---- (1986). 'On Being a Linguistic Anthropologist', *Annual Review of Anthropology* 15: 1-24.

Grinevald, C. (2000). 'A Morphosyntactic Typology of Classifiers', in G. Senft (ed.), *Systems of Nominal Classification*, 50-92. Cambridge: Cambridge University Press.

---- (2002). 'Making Sense of Nominal Classification Systems: Noun Classifiers and the Grammaticalization Variable', in I. Wischer and G. Diewald (eds.), *New Reflections on Grammaticalization*, 259-275. Amsterdam: John Benjamins.

---- and Seifart, F. (2004). 'Noun Classes in African and Amazonian Languages: Towards a Comparison', *Linguistic Typology* 8: 243-285.

Hajek, J. (2007). 'Sound Systems of the Asia-Pacific: Some Basic Typological Observations', in P. Austin and A. Simpson (eds.), *Endangered Languages*, 206-222. Hamburg: Helmut Buske Verlag.

Hale, K., Krauss, M., Watahomigie, L. J., Yamamoto, A. Y., Craig, C., Jeanne, L. M., and England, N. C. (1992). 'Endangered Languages', *Language* 68: 1-42.

Haspelmath M. (1999). 'Optimality and Diachronic Adaptation', *Z. Sprachwiss*. 19:180-205.

---- (2004). 'Explaining the Ditransitive Person-Role Constraint: A Usage-Based Account', *Constructions* 2: 1-71.

---- (2007). 'Pre-established Categories Don't Exist: Consequences for Language Description and Typology', *Linguistic Typology* 11: 119-132.

---- (2008). 'Comparative Concepts and Descriptive Categories in Cross-linguistic Studies,' manuscript, Max Planck Institute for Evolutionary Typology, Leipzig. Online at http://www.eva.mpg.de/lingua/staff/haspelmath/papers.php.

Hill, J. H. (2006). 'Writing Culture in Grammar in the Americanist Tradition', in F. Ameka, A. Dench, and N. Evans (eds.), *Catching Language: The Standing Challenge of Grammar Writing*, 609-628. Berlin: Mouton de Gruyter.

Himmelmann, N. (1998). 'Documentary and Descriptive Linguistics', *Linguistics* 36: 161-95.

---- and Evans, N. (2007). 'Workshop proposal: Linguistic Typology and Language Documentation,' 24-25 September, Paris, France.

Hockett, C. (1993). 'Direction in the Algonquian Verb: A Correction,' in *The Rice Papers*, first installment, June 1993. Manuscript, Rice University.

Hudson, R. (2007). 'Towards a Useful Theory of Language', in P. Austin, O. Bond and D. Nathan (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory*, 3-11. London: SOAS. Online at www.hrelp.org/eprints/ldlt_02.pdf.

Jakobson, R. (1960). 'Closing Statement: Linguistics and Poetics', in T. Sebeok (ed.), *Style in Language*, 350-377. Cambridge, Mass: MIT Press.

Johnson, K. (2002). 'Review of *Vowels and Consonants* by Peter Ladefoged', *Phonology* 19: 1-5.

Keller, R. (1994). *On Language Change: The Invisible Hand in Language*. London: Routledge.

Khanina, O. (2009). 'The Symbiosis of Descriptive Linguistics and Typology: A Case Study of Desideratives', in P. Epps and A. Arkhipov (eds.), *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, 199-221. Berlin: Mouton de Gruyter.

Lehmann, C. (2001). 'Language Documentation: A Program,' in Walter Bisang (ed.), *Aspects of Typology and Universals*. Deutsche Gesellschaft für Sprachwissenschaft: Akademie Verlag.

Lidz, L. (2007). 'Combinatorial Complexity: An Example from Yongning Na (Mosuo)', paper presented at the Pre-ALT workshop on Linguistic Typology and Language Documentation, 24-25 September, Paris, France.

Lüpke, F. (2007). 'Vanishing Voice – The Morphologically Zero-coded Passive of Jalonke', in P. Austin and A. Simpson (eds.), *Endangered Languages*, 173-190. Hamburg: Helmut Buske Verlag.

Michael, L. (2008). *Nanti evidential practice: language, knowledge, and social action in Amazonian society*, Ph.D. dissertation, University of Texas at Austin.

Moravcsik, E. (2007). 'What is Universal About Typology?' *Linguistic Typology* 11: 27-41.

Munro, P. (2003). 'Field Linguistics', in M. Aronoff and J. Rees-Miller (eds.), *The Handbook of Linguistics*, 130-149. Oxford: Blackwell Press.

Nettle, D. and Romaine, S. (2000). *Vanishing Voices: The Extinction of the World's Languages*. Oxford: Oxford University Press.

Nichols, J. (2007). 'What, if Anything, is Typology?' *Linguistic Typology* 11: 231-238.

O'Connor, L. (2009). 'All Typologies Leak: Predicates of Change in Lowland Chontal of Oaxaca', in P. Epps and A. Arkhipov (eds.), *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, 343-364. Berlin: Mouton de Gruyter.

Olawsky, K. (2007). 'ObViouS OVS in Urarina Syntax,' in P. Austin and A. Simpson (eds.), *Endangered Languages*, 45-72. Hamburg: Helmut Buske Verlag.

Ospina Bozzi, A. (2002). *Les structures élémentaires du Yuhup Maku, langue de l'Amazonie Colombienne: Morphologie et syntaxe*, Ph.D. diss., Université Paris 7 – Denis Diderot.

Plank, F. (ed.). (2003). 'Das grammatische Raritätenkabinett', electronic database, http://typo.uni-konstanz.de/rara/intro/.

---- (2007). 'Extent and Limits of Linguistic Diversity as the Remit of Typology – But Through Constraints on WHAT is Diversity Limited?' *Linguistic Typology* 11: 43-68.

---- and Filimonova, E. (2000). 'The Universals Archive: A Brief Introduction to Prospective Users', *Sprachtypologie und Universalienforschung* 53: 109-123.

---- and Mayer, T. (2006). 'The Universals Archive Goes Do-It-Yourself', *Linguistic Typology* 10: 457-459.

Rivet, P., Kok, P., and Tastevin, C. (1925). 'Nouvelle Contribution à l'Étude de la Langue Makú', *International Journal of American Linguistics* 3: 133-192.

Rodrigues, A. D. (2000). 'Panorama das Línguas Indígenas da Amazônia', in F. Quiexalós and O. Renault-Lescure (eds.), *As Línguas Amazônicas Hoje*, 15-28. São Paulo: IRD/ISA/MPEG.

Rumsey, A. (2007). 'Typology, Meaning and Language Documentation: An Anthropological Perspective', paper presented at the Pre-ALT workshop on Linguistic Typology and Language Documentation, 24-25 September, Paris, France.

Sapir, E. (1921). *Language.* New York: Harcourt, Brace, and Co.

Schultze-Berndt, E. (2007). 'On Manners and Paths of Refining Talmy's Typology of Motion Expressions via Language Documentation', in P. Austin, O. Bond and D. Nathan (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory*, 223-233. London: SOAS. Online at www.hrelp.org/eprints /ldlt_02.pdf

Seifart, F. (2009). 'Multidimensional Typology and Miraña Class Markers', in P. Epps and A. Arkhipov (eds.), *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, 365-387. Berlin: Mouton de Gruyter.

Seki, L. (2000). *Gramática do Kamaiurá, língua Tupi-Guarani do Alto Xingu*. Campinas: Editora UNICAMP.

Sherzer, J. (1990). *Verbal Art in San Bias: Kuna Culture Through its Discourse.* Cambridge: Cambridge University Press.

Shopen, T. (ed.). (2007). *Language Typology and Syntactic Description*, 2nd edition. Cambridge: Cambridge University Press.

Song, J. J. (2007). 'What or Where Can We Do Better? Some Personal Reflections on (the Tenth Anniversary of) Linguistic Typology', *Linguistic Typology* 11: 5-22.

Talmy, L. (1985). 'Lexicalization Patterns: Semantic Structure in Lexical Forms', in T. Shopen (ed.), *Language Typology and Syntactic Description*, vol. 3, 36-149. Cambridge: Cambridge University Press.

---- (1991). 'Path to Realization - via Aspect and Result: A Typology of Event Conflation', *Berkeley Linguistics Society* 17: 480-519.

Thurston, W. (2007). 'Describing Languages and Ethnographic Fieldwork', in J. Siegel, J. Lynch, and D. Eades (eds.), *Language Description: History and Development*, 15-26. Amsterdam: John Benjamins.

Van Valin, R. (2007). 'Some Thoughts on the Reason for the Lesser Status of Typology in the USA as Opposed to Europe', *Linguistic Typology* 11: 253-257.

Wälchli, B. (2007). 'Potential Challenges of Typologies Based on Primary Data, Illustrated With a Case Study on Nominal Plural', paper presented at the Pre-ALT workshop on Linguistic Typology and Language Documentation, 24-25 September, Paris, France.

Woodbury, A. (2003). 'Defining Documentary Linguistics', in P. Austin (ed.), *Language Description and Documentation*, vol. 1, 35-51. (Endangered Languages Project.) London: School of Oriental and African Studies.

[1] As Woodbury (2003: 42) puts it, "there is a dialectical relationship between corpus and apparatus – the corpus informs the analytic apparatus; but analysis – including everything you bring to the table when doing grammatical and lexical elicitation – in turn also informs the corpus. Likewise, almost any presentation of documentary work requires grammatical analysis – transcription requires a phonological analysis, and lexical presentation in the form of a thesaurus or dictionary requires morphological and lexical analysis." The view taken here contrasts to some degree with that of Himmelmann (1998), who posits a sharper break between documentation and description. Both documentation and description are sometimes referred to as 'field linguistics'.

[2] For example, Amazonia has some 240 languages (still spoken), comprising 52 distinct linguistic groupings (Rodrigues 2000).

[3] This distinction between the language-specific and the general is aided by the convention of capitalizing terms applied to language-specific categories, e.g. 'Perfective aspect', in order to distinguish these from the corresponding cross-linguistic categorial notions (e.g. Comrie 1976, Bybee 1985).