The typological database of the *World Atlas of Language Structures*

MARTIN HASPELMATH

1. Introduction

The *World Atlas of Language Structures* (often abbreviated as *WALS*) is primarily a book with 142 world maps showing the global distribution of structural features of language. It was put together by Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie at the Max Planck Institute for Evolutionary Anthropology between 1999 and 2004, and published by Oxford University Press in July 2005 (Haspelmath et al. 2005). Over forty authors contributed to it, each structural feature (and thus each map) being the responsibility of a single author or team of authors. A sample map is shown in Figure 1.

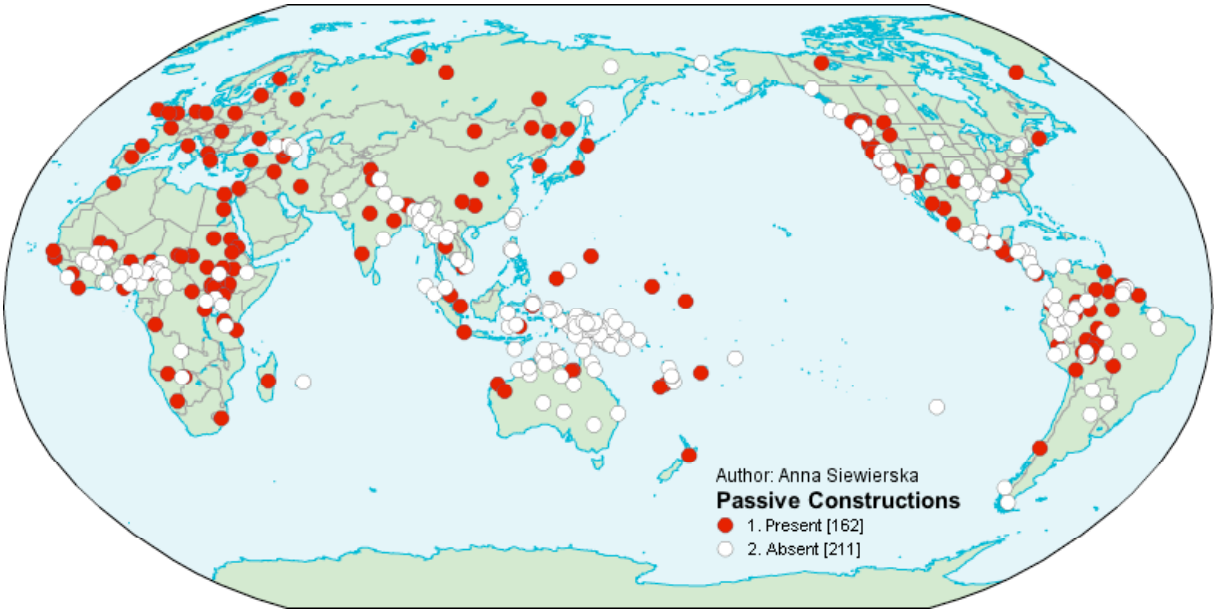

Figure 1. The *WALS* map "Passive Constructions", by Anna Siewierska (Siewierska 2005)

On the maps, each language is shown by a dot (most often a circle), and different colours stand for different structural types (or feature values). Thus, in the map in Figure 1 the white dots are languages lacking a passive constructions, and the red dots are languages with a passive construction. The *World Atlas of Language Structures* thus resembles a traditional dialect atlas, but the coding points are not places that the authors actually visited. Instead, they stand for languages on which the authors obtained information through published descriptions (reference grammars, dictionaries, scholarly articles, but occasionally also personal communications from experts and/or speakers of the language). Only at most 10% of the world's languages can be said to have been described reasonably well, so the maps only show about 400 languages on average (out of the 6000-7000 languages that were still spoken at the end of the 20th century). The features have at least two different values, and at most nine, because more than nine different colours (or colour-shape combinations) are difficult to distinguish visually on a map.

The editors' task thus consisted in assembling a database from the authors that primarily consisted of one two-column table for each feature, giving pairs of language names and feature values. In addition we asked for bibliographical references and page numbers. A very partial sample table (showing just five languages) is shown in Table 1.

language name	feature value (1:present, 2: absent)	author-year	pages
Apurinã	1:present	Facundes 2000	522
Arapesh	2:absent	Conrad and Wogiga 1991	14
Evenki	1:present	Nedjalkov 1997	217-225
Koasati	1:present	Kimball 1991	138
Tunica	2:absent	Swanton 1921	5-21

Table 1. A very partial table exemplifying the data for Figure 1.

In addition, the authors were asked to provide a 2000-word text giving a description of the feature and providing full definitions of the various values. These texts are printed in the atlas on the two pages preceding the two map pages.

Since linguists may want to use the data underlying the *World Atlas of Language Structures* in a variety of ways, an electronic version of the atlas was published together with the book on a CD-ROM: The Interactive Reference Tool, programmed by Hans-Jörg Bibiko at the Max Planck Institute for Evolutionary Anthropology. This programme allows users to view the maps of the printed atlas and to display the data in a variety of ways, to conduct automatic searches, to export data and maps, and to create compound features based on the standard 141 features of the printed version.

Thus, the WALS data can be seen as a single complex database, consisting of 141 databases on structural features that are linked by a common metadata scheme (data on languages, references, and so on).

2. Research questions

There were two main research questions that the editors and the authors wanted to address:

(i) What correlations exist between structural features in different areas of grammar? For example, is it true that languages with little verb inflection tend not to make a past-nonpast distinction (not even a noninflectional one)? Is it true that languages with large vowel inventories tend to have small consonant inventories, and vice versa?

(ii) What geographical patterns are exhibited by the structural features? For example, is it true that tone distinctions are found primarily in sub-Saharan Africa and Southeast Asia? (The geographical perspective on the distribution of structural features is generally called *areal typology*.)

That interesting correlations between different structural features can be found has been well-known since Greenberg (1963), and since the 1980s the search for correlations has also been prominent in generative syntax. Since the 1970s a substantial amount of systematic large-scale cross-linguistic research (i.e. research involving 50 or more languages from all areas of the world) has been carried out, and it was obviously desirable to put the typological data together in such a way that potential correlations can be tested easily and automatically. Thus, the editors approached linguists who they know had gathered data from a large number of languages and asked them to contribute one or several chapters to *WALS*. We also enlisted several doctoral students who were in the process of gathering data for their

dissertations, and a number of typologists only started gathering data on a large-scale basis when they heard about the project in 1999 and 2000. (We did not try to incorporate any of the early published work from the 1970s and 1980s whose authors were no longer actively involved in typological research.) *WALS* includes data from three projects that are described in more detail in this book: The StressTyp database (Rob Goedmans and Harry van der Hulst), the Surrey Morphology Group's syncretism database (Matthew Baerman and Dunstan Brown), and the database on intensifiers and reflexives at the Freie Universität Berlin (Ekkehard König and associates).

That structural features tend to cluster geographically has also been known for quite a while. Sprachbund phenomena have been discussed for a number of areas in various parts of the world, and already Jakobson (1931) ventured the hypothesis of a Eurasian Sprachbund based on a few phonological features. But the issue of large areal patterns became prominent in language typology only with the publication of Dryer (1989) and Nichols (1992). Especially the latter included detailed discussions of areal patterns, but contained virtually no maps. Areal typology within Europe received a boost from van der Auwera (1998) and related work in the EUROTYP project, which showed that even outside the well-known Balkan area, many geographical patterns can be found (see also Haspelmath 2001). So at the end of the 1990s the time seemed ripe for a larger enterprise that put much of the available (and also a lot of new) cross-linguistic data on maps on a global scale.

Thus, *WALS* tried to address two goals simultaneously that are not logically linked to each other. The search for correlations can proceed without any geographical information, and the search for areal patterns need not be concerned with correlations. However, combining the two goals had a number of clear benefits:

First, the stated goal of publishing an atlas helped motivate those contributors that were primarily interested in correlations, because their contribution was published in the form of a "chapter" consisting of two text pages (written by them) and two map pages (for which they provided the underlying data). Four pages is not much on a CV, but the atlas chapters are conventional publications that can be cited easily, and the authors can get credit for their work in this way. If we had just published an electronic database without a book, the authors would probably have been much more reluctant to share their data. (And if we had proposed publishing the data as printed tables in a book, we would not have found a publisher, or the book would not have been read.)

Second, if we had just focused on the areal patterns without taking correlations into account, we might have limited ourselves to a printed atlas. But the goal of finding correlations forced us to include a way of allowing the user to search for correlations in a number of ways. As a result even those chapters that were primarily included for their geographical-pattern interest can now be used for finding correlations.

Third, at least since Dryer (1989) it has been widely known that finding valid correlations presupposes some awareness of geographical patterns, just as it presupposes awareness of genealogical patterns. For example, if we limit ourselves to the languages of Africa and Europe, it seems as if the presence of a tone distinction precludes the presence of a rounding distinction in front vowels (i.e. *i* vs. *ü*, *e* vs. *ö*). African languages tend to show tone, European languages front vowel rounding. But the world-wide picture is rather different: Tone contrasts are found especially in Africa and Southeast Asia, while front rounded vowels are found especially in northern Eurasia (cf. Figure 2 below). In China, there are not fewer languages with front rounded vowels than in Europe, and all of them have a tone distinction.

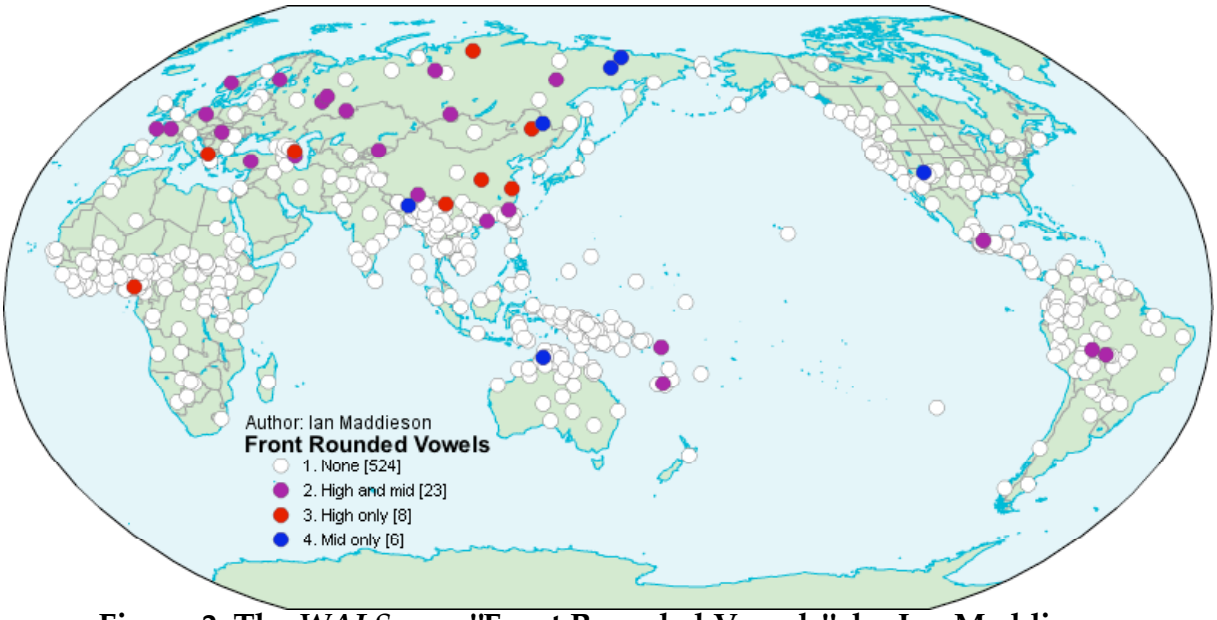

Figure 2. The *WALS* map "Front Rounded Vowels", by Ian Maddieson (Maddieson 2005a)

Thus, the correlation goal and the areality goal fit together very well, and in actual fact nowadays most comparative linguists have both research questions in mind when they study a particular phenomenon in a large number of languages.

3. Database design

The *WALS* database consists of three main tables: The Data table, the Features table, the Languages table. (In addition, there are other metadata tables such as the references table that I will not talk about here.) As our primary goal was to give a representative picture of the world's linguistic diversity, we just asked for minimal information on each language-feature pair: a feature value for the dot colour/shape on the map, and references including page numbers (cf. Table 1 above). In addition, we allowed the authors to provide an example (since this is very time-consuming, only a fairly small number of features include examples). We also allowed up to five references, so we had to include five author-year and page number fields in the data table. Thus, the Data table ended up being more complicated than the simple Table 1 above. A list of the fields is given in Table 2, with three sample records. In this table, the Language Number field and the Feature Number field are necessary to relate the Data table to the Languages and Features tables, respectively. The data are from Siewierska 2005, Gil 2005, and Corbett 2005.

Language Number:	1233 (=Apurinã)	645 (=Nauruan)	2011 (=Lak)
Feature Number:	107 (=Passive)	55 (=Classifiers)	30 (=Number of Genders)
Value:	1 (=present)	3 (=obligatory)	4 (=four)
Example:	-	-	-
Author-year 1:	Facundes 2000	Kayser 1993	Corbett 1991
Page Numbers 1:	522	68-76	24-26
Author-year 2:	-	Lynch 1998	Xajdakov 1980
Page Numbers 2:	-	120	204-213
Author-year 3:	-	-	Murkelinskij 1967
Page Numbers 3:	-	-	166-167
...			

Table 2: The Data table of the *WALS* database: Three sample records

The Features table primarily contains the feature name, the feature number, and the names of the feature values:

Feature Number:	30	33	105
Feature Name:	Number of Genders	Coding of Nominal Plurality	Ditransitive Constructions: The Verb 'Give'
Value Name 1:	None	Plural prefix	Indirect-object construction
Value Name 2:	Two	Plural suffix	Double-object construction
Value Name 3:	Three	Plural stem change	Secondary-object construction
Value Name 4:	Four	Plural tone	Mixed
Value Name 5:	Five or more	Plural reduplication	-
Value Name 6:	-	Mixed plural	-
Value Name 7:	-	Plural word	-
Value Name 8:	-	Plural clitic	-
Value Name 9:	-	No plural	-

Table 3: The Features table of the *WALS* database: Three sample records (Corbett 2005, Dryer 2005, Haspelmath 2005)

The actual database is a little more complex than shown in Table 3: For all feature values, we actually have long and short value names. The long names appear in the text, and the short names appear on the map legend and in the electronic version. Moreover, we added information about dot colours/shapes in nine additional Value Colour fields.

The Languages table crucially includes the language number, the language name, the location, and the *WALS* code (i.e. the three-letter abbreviation that is shown on each dot on the printed maps). Identifying the languages that the authors provided information on proved to be a very time-consuming task, since language names are often not sufficient to identify a language. Only after *WALS* was completed was an ISO standard for unique identification of languages established (ISO 639-3), and the discipline of linguistics is still far from accepting such a standard. It will take a while before linguistics publications include a unique language identifier as a matter of course. Thus, the WALS authors were required only to give a language name and their sources, and where there was a problem (e.g. in identifying the right dialect of Quechua or Berber), the editors consulted the sources in order to establish the identity of the language. Where alternative names exist in the literature, the editors tried to choose the name form that is currently the most common among linguists and is the most acceptable to the speakers.

In addition, in order to facilitate the identification of the languages, the editors added Ethnologue 14 and Ethnologue 15 codes (the latter being largely identical with

ISO 639-3), the names from Ethnologue, and the names chosen in the two other major published language catalogues, Ruhlen (1987) and Moseley & Asher (eds.) (1994). Moreover, for each language its family and genus was determined, and for some larger families several subfamilies are distinguished (e.g. Chadic within Afro-Asiatic, or Munda within Austro-Asiatic). Finally, for each language we have information on the country (or countries) where it is (primarily) spoken. Thus, the Languages table contains the 13 fields shown and exemplified in Table 4.

Language Number:	645	1038	2205
Language Name:	Nauruan	Nuuchahnulth	Dâw
Location:	166°55E 0°30S	126°40W 49°40N	67°05W 0°15S
WALS Code:	nau	nuu	daw
Ethnologue 14 code:	NRU	NOO	KWA
Ethnologue 15 code:	nau	noo	kwa
Ethnologue name:	Nauruan	Nootka	Kamã
Ruhlen name:	Nauruan	Nootka	-
Asher & Mosely:	Nauruan	Nootka	Kamán
Family:	Austronesian	Wakashan	Vaupés-Japurá
Subfamily:	Eastern Malayo-Polynesian	-	-
Genus:	Oceanic	Southern Wakashan	Vaupés-Japurá
Country:	Nauru	Canada	Brazil

Table 4: The Languages table of the *WALS* database: Three sample records

Of course, it would have been desirable to include other sociolinguistic information of various sorts, such as the number of speakers, the use of the language in the media and in schools, the amount of bili ngualism, and so on. And given that the database is part of an "atlas", it is natural to ask for a more precise indication of the place(s) where the language is spoken, e.g. in the form of polygons (rather than just a single dot at the centre of the area where the language is spoken). Unfortunately, information of this sort is available only for a small percentage of the languages, so we did not try to include it. Moreover, on the atlas maps we did not want to privilege languages with a large number of speakers, or languages that are spoken over a wide area, so even if polygon information had been available, we would not have chosen it as the primary means of presenting the data on maps.

4. Implementation

The database described in the preceding section was implemented in FileMaker Pro, and since the use of database software was (at least until recently) not universal even among typologists, often text files had to be converted into the right database format.

The database was published on a CD-ROM accompanying the printed atlas together with a programme (the Interactive Reference Tool) that allows users to display the data in a variety of ways, to conduct automatic searches, to export data and maps, and to create compound features based on the standard 141 features of the printed version.

Users of the electronic database can customize the map in various ways: show major cities and country names, remove country boundaries and rivers, and replace the light green/light blue base map by a topographic map showing altitude levels. The language dots can be shown in five different sizes, and the language name can be shown either as three-letter *WALS* code inside the symbol, or in full to the right of the symbol. The colors and shapes of the symbols can be changed. When the mouse

pointer moves over the dot, the full name is shown, and when clicking on a dot, a window with further information on the language opens (including the data source). Users can also zoom in on areas with high dot density, closely enough to see all dots separately, and drag on a map to see adjacent areas. Maps can be exported and printed, and various user-defined selections can be saved for future use.

The Interactive Reference Tool allows users to manipulate the standard features in two ways: values can be removed (if they are not of interest in a certain context), and several values can be merged into a single value. For instance, the five values of chapter 1 (small, moderately small, average, moderately large, large) can be reduced to three (below average, average, above average) with just two mouse clicks.

Users can search for language names, genus names, family names, country names, and even for text in bibliographic entries. It is possible, for instance, to find and display all languages beginning with *X*, all languages belonging to the Austronesian family, all languages spoken in Colombia, or all languages described by Jeffrey Heath. On the maps that only show languages (without giving information about the features), different dot colours may stand for different families or different genera.

The Interactive Reference Tool contains no particular reseources for assessing areal patterns beyond its map-making capability. Whether an apparent geographical clustering of a particular type is significant or just looks significant will have to be decided in different ways. (So far the methodology of assessing areality is a very underdeveloped area in comparative linguistiuics.) (See Comrie 2006+ for more on WALS as a research resource for areal typology.)

However, the Tool was designed to help the comparative linguist find correlations between different features, genealogical information, and geographical information. The most frequent question of theoretical linguists is perhaps whether two features corelate. To test this, users can create their own compound features. For example, they may want to know whether the existence of tone in a language is correlated with the type of syllable structure. Both features have three values (tone: none, simple, complex; syllable structure: simple, moderately complex, complex), so by combining them, one gets nine possible values, as shown in Table 5.

Combined value	languages
No tones AND Simple syllable structure	28
No tones AND Moderately complex syllable structure	135
No tones AND Complex syllable structure	112
Simple tone system AND Simple syllable structure	21
Simple tone system AND Moderately complex syllable structure	75
Simple tone system AND Complex syllable structure	23
Complex tone system AND Simple syllable structure	11
Complex tone system AND Moderately complex syllable structure	58
Complex tone system AND Complex syllable structure	8

Table 5. Result of combining two three-valued *WALS* features

The programme automatically creates a compound feature with these nine values, shows the number of languages in each value, suggests a symbol for each value, and displays a map of the compound feature. More complex ways of creating compound features are also possible and are described in detail in the Manual of the programme.

5. Limitations

The World Atlas of Language Structures was conceived of as a five-year project and primarily as an atlas, and as a result there are a number of limitations of the database that from some perspectives leave certain things to be desired. Especially if one tries to approach the *WALS* database from a quantitative/statistical point of view, one quickly realizes that *WALS* is not perfect.

Since the feature value are the core piece of information provided by *WALS*, the entire database can be seen as primarily consisting of a 141-by-2560 matrix (141 features, 2560 languages) with 360,960 cells that can be filled with an integer between 1 and 9 (the feature value). However, on average each map shows only around 400 languages, so that there are only about 58,000 cells filled with data points—about 84% of the cells are empty. Thus, although *WALS* makes a huge amount of information readily available for the first time and is therefore widely regarded as a milestone in the history of comparative linguistics, from a statistician's point of view a problem is "the large amount of missing data" (Cysouw et al. 2007). This limitation can be overcome only by gathering further data, the most expensive part of the entire enterprise. For most of the gaps, original fieldwork with the speakers of the languages would be required. The alternative option of limiting the admitted languages strictly to those that have a coiding for all the features, or of limiting the admitted features to those that have a coding for all the languages, would have entailed discarding a large amount of information that might be invaluable from other perspectives, so it was not seriously considered.

A limitation that was imposed by the goal of making an atlas is the restriction of the feature values to nine. While the most salient structural parameters of grammatical structure do not normally have more than a handful of values (e.g. with two values: configurational vs. nonconfigurational; with three values: head-final vs. head-initial vs. free order; with four values: head-marking vs. dependent-marking vs. double marking vs. zero marking), it is easy to define a parameter in such a way that there are more than nine values. For example, chapter 33 on the coding of nominal plurality (Dryer 2005) distinguishes plural prefixes from plural suffixes (cf. Table 3 above), but it has "plural clitic" and "plural word" as unitary types without differentiating between proclitics and enclitics, or between preposed and postposed plural words. Another striking case concerns features where elements are counted, as in chapter 30 (on the number of genders, Corbett 2005; cf. Table 3 above). The fifth value "five or more" lumps together a potentially large number of different situations, and these could have been distinguished. What we could have done (and what a future more sophisticated project of this sort will probably do) in such situations is to include two levels of feature values: On the one hand, a level of feature-value detail where a large number of distinct types are distinguished in the database to capture a maximum of information. Since this information is difficult to display on a map (at least if the map is meant to be interpreted directly by human observers), similar types could then be lumped together exclusively for the purposes of map representation. In *WALS*, the decision was taken to maintain strict identity between the database and the maps to simplify the procedure, but this is of course not necessary.

A consequence of the upper limit on values is that many maps work with a relatively uninformative value "other" or "mixed". For example, in map 105 on ditransitive constructions (Haspelmath 2005; cf. Table 3 above) the "mixed" type can be a mixture of type 1 and 2, of type 2 and 3, of type 1 and 3, or of all three types. These different mixtures would not have been easy to show transparently on the map, so it was decided to lump them together in a single mixed type. Such "mixed" or "other" values are problematic for statistical analyses, especially similarity

analyses, because from the fact that two languages are coded as "other" one cannot conclude that they are more similar to each other than to any of the other types. So again a more sophisticated future project would distinguish the various mixed types at the database level, and would lump them together at the level of map representation.

Another decision that was driven by the goal of making maps was to display several unrelated features on a single map, and to treat them as a single feature in the database. For example, in ch. 9 ("Presence of uncommon consonants", Maddieson 2005b), seven values are displayed on the map:

value	number of languages
1. None of the four uncommon consonants	448
2. Presence of clicks	9
3. Presence of labial-velars	45
4. Presence of pharyngeals	21
5. Presence of "th" sounds	40
6. Presence of clicks, pharyngeals and "th" sounds	1
7. Presence of pharyngeals and "th" sounds	2

Table 6: The seven values distinguished on map 9 (Maddieson 2005b)

But this seven-valued feature is of course just a conflation of four binary features, concerning the presence or absence of clicks, labial-verlars, pharyngeals and "th" sounds. These are all fairly rare sounds, so maps showing their distribution would be relatively uninteresting, because almost all of the languages would have the value "absence". Even with four rare consonants, 79% of the languages have the value "None" and appear as white dots on map 9. A future version of the database should distinguish the underlying single features from the composite features that human users like to see on maps. This will lead to a proliferation of features, because a large number of the *WALS* features can be decomposed further. Consider map 33 on the "Coding of nominal plurality" (Dryer 2005), whose WALS values are shown as ean example in Table 3 above. The nine values are repeated in Table 7.

value	number of languages
1. Plural prefix	118
2. Plural suffix	495
3. Plural stem change	5
4. Plural tone	2
5. Plural reduplication	8
6. Mixed plural	34
7. Plural word	150
8. Plural clitic	59
9. No plural	86

Table 7: The nine values distinguished on map 33 (Dryer 2005)

These nine values can be recoded in the following way as primitive features:

a. presence/absence of nominal plurality (1-8/9)
b. fusion of plural coding: word/clitic/affix/stem change (7/8/1-2/3-4)
c. position of plural element: preceding/following (1/2)
d. type of stem change: segmental/tonal (3/4)
e. source of plural element: reduplicated/specified (5/1-2, 7-8)

Such features are less interesting for the human interpreter and data provider, but

from a database point of view they are more straightforward than *WALS*'s composite features.

Another limitation that is highlighted by Cysouw et al. (2005) is the fact that the concepts used in the feature and value descriptions are not standardized. Terms like "case" or "clitic" may not have exactly the same meaning across different chapters. In other words, the WALS chapters are not based on a standard ontology, but are in conformity with the normal practice of linguistics: Technical concepts have to be defined anew by each author, because there is not enough common ground among linguists even for fairly basic concepts. This is a limitation that will be much harder to overcome than the others mentioned in this section.

6. Prospects

Clearly, the next steps to be taken by the editors and other interested comparative linguists must aim to expand the *WALS* database and to overcome some of the limitations mentioned in the last section. Of course, this will be a long-term process, but now that *WALS* exists many of the challenges that are ahead of us have come into sharper focus.

A relatively straightforward improvement in access to the database would be its free availability on the web, and it is hoped that this will be achieveable within a few years. In fact, it had been suggested from the beginning by some of the contributors that a *WALS* on the web was sufficient, and that a book was not eeded. However, since the discipline does not yet have a standard way of giving credit to scholars who supply their data (plus data description) to a larger electronic database, publishing a book was probably the right decision in 1999. It is clear, though, that it is an important task for linguists (and scientists more generally) to develop conventions for giving recognition to scholars who contribute their datasets to a publicly accessible database.

Thus, the current plan is not to publish a second volume of *WALS*, but to transform it into a freely available, extendable, constantly growing and improving web resource.

References

Comrie, Bernard. 2006+. "Areal typology and the *World Atlas of Language Structures.*" To appear in a volume from a Thessaloniki conference.

Corbett, Greville C. 2005. "Number of genders." In: *WALS*, 126-129.

Cysouw, Michael & Jeff Good & Mihai Albu & Hans–Jöerg Bibiko. 2005. "Can GOLD "cope" with WALS? Retrofitting an ontology onto the World Atlas of Language Structures." In: *Proceedings of the 2005 E-MELD workshop* (http://emeld.org/workshop/2005/proceeding.html)

Cysouw, Michael & Mihai Albu & Andreas Dress. 2007. "Analyzing feature consistency using dissimilarity matrices." *Sprachtypologie & Universalienforschung*

Dryer, Matthew S. 1989. "Large linguistic areas and language sampling." *Studies in Language* 13:257-92.

Dryer, Matthew S. 2005. "Coding of nominal plurality." In: *WALS*, 138-141.

Gil, David. 2005. "Numeral classifiers." In: *WALS*, 226-229.

Greenberg, Joseph H. 1963. "Some universals of grammar with particular reference to the order of meaningful elements." In: Greenberg, Joseph H. (eds.) *Universals of grammar*, 73-113. Cambridge, Mass.: MIT Press.

Haspelmath, Martin. 2001. "The European linguistic area: Standard Average European." In: Haspelmath, Martin & König, Ekkehard & Oesterreicher, Wulf & Raible, Wolfgang (eds.) *Language typology and language universals.* (Handbücher zur Sprach- und Kommunikationswissenschaft) Berlin: de Gruyter, 1492-1510.

Haspelmath, Martin. 2005. "Ditransitive constructions: The verb 'give'." In: *WALS*, 426-429.

Haspelmath, Martin & Matthew S. Dryer & David Gil & Bernard Comrie (eds.) 2005. *The World Atlas of Language Structures.* Oxford: Oxford University Press.

Jakobson, Roman. 1931. Über die phonologischen Sprachbünde. *Travaux du Cercle Linguistique de Prague* 4.234-40. (Réunion phonologique internationale tenue a Prague, 18-21/XII, 1930.) (Reprinted, 1971: *Roman Jakobson. Selected Writings*, vol. 1, 137-43. The Hague: Mouton.)

Maddieson, Ian. 2005a. "Front rounded vowels." In: *WALS*, 50-53.

Maddieson, Ian. 2005b. "Presence of uncommon consonants." In: *WALS*, 82-85.

Moseley, Christopher and R.E. Asher (eds.) 1994. *Atlas of the World's Languages*. London: Routledge.

Nichols, Johanna. 1992. *Linguistic diversity in space and time.* Chicago: The University of Chicago Press.

Ruhlen, Merritt. 1987. *A Guide to the World's Languages 1: Classification.* Stanford: Stanford University Press.

Siewierska, Anna. 2005. "Passive constructions." In: *WALS*, 434-437.

van der Auwera, Johan (1998), "Conclusion", In: van der Auwera, Johan (ed.), *Adverbial constructions in the languages of Europe*, Berlin: Mouton de Gruyter, 813-36.

WALS: Haspelmath et al. 2005.