

Tokyo LingDy Grammar Workshop: Position Paper Notes

Grammar-writing: looking into the future

Nicholas Evans, Australian National University
nicholas.evans@anu.edu.au

1. Grammar-writing: Ideals and prospects

1.1 Grammars of the future

- hypertext, with
 - direct clickable links to primary data (e.g. Thieberger 2006, Morey 2005)
 - multi-level descriptions to enable the reader to engage at different levels of detail
 - cross-referencing navigation
 - [a first attempt at this is Sebastian Nordhoff's grammar of Sri Lankan Malay, downloadable at <http://www.lotpublications.nl/publish/issues/Nordhoff/index.html>]
 - direct links to global ontology of linguistic categories so as to compare language-specific descriptive terms to those used cross-linguistically
 - many pathways into grammar for typological queries (cf Zaefferer 2006)

1.2 Grammar-writing within the broader process of documentation and description

- the need for quality and discovery
 - 'Linguistic theory will never be moved ahead as far by answers to questions we already know enough to ask as it will by discoveries of the unexpected' (Mithun 2001:45)
- typological importance of clade leaders (i.e. descriptions of undocumented clades) for maximal ontology
- but importance of multi-member descriptions for
 - cross-checking / sharpening / balancing of accounts
 - intra-genetic typology (e.g. for testing hypotheses about covariation)
 - we can't assume that the most interesting phenomena will automatically turn up in a single sample per clade (e.g. Kayardild and modal or verbal case (Evans 1995) – not in Yukulta, the first-described Tangkic language)
 - understanding of historical pathways, variation and change

2 Grammar-writing and linguistic diversity: a challenge to world science

2.1 Grammar-writing and the negative correlation between linguistic diversity and research infrastructure at national levels

- extremely unbalanced situation world-wide
 - by language family (cf Australian vs. Papuan; Mayan vs. Mixe-Zoquean)
 - by 'training area' (Netherlands vs. UK; Australia vs. USA; Mexico vs. Pakistan)
- Most of the top mega-diverse countries lack good training or output

Countries in the Top 25 megadiverse countries (by languages, lineages or both) lacking proper descriptive training (defined as: no first-class description by a linguist partly or primarily trained in that country)

PNG (1; 2)	Indonesia (2;3)
Nigeria (3; 20)	Cameroon (7; 23)
Zaire (9)	Vanuatu (12)
Tanzania (13, 22)	Sudan (14; 10)
Malaysia (15)	

Countries with some linguistic training but not producing grammatical descriptions at an appropriate level for the amount of linguistic diversity

Mexico	India	China	Ethiopia
--------	-------	-------	----------

<i>Rank</i>	<i>Endemic languages</i>		<i>Endemic linguistic lineages</i>	
1	PNG	847	USA	64
2	Indonesia	655	PNG	58
3	Nigeria	376	Indonesia	37
4	India	309	Brazil	31
5	Australia	269	Mexico	24
6	Mexico	230	Colombia	24
7	Cameroon	201	Australia	22
8	Brazil	185	Peru	21
9	Zaire	158	Russia	17
10	Philippines	153	Sudan	15
11	USA	143	Canada	14
12	Vanuatu	105	Bolivia	13
13	Tanzania	101	Venezuela	11
14	Sudan	97	India	10
15	Malaysia	92	China	9
16	Ethiopia	90	Ethiopia	8
17	China	77	Chad	8
18	Peru	75	Argentina	8
19	Chad	74	Ecuador	8
20	Russia	71	Nigeria	6
21	Solomon Is.	69	Burkina Faso	6
22	Nepal	68	Tanzania	5
			Cameroon	5
			Georgia	5
			Chile	5
			Laos	5
23	Colombia	55		
24	Côte d'Ivoire	51		
25	Canada	47		

Table 1.2. The top 25 megadiverse countries, for two measures of language diversity. Left column from Harmon 1996; right column calculated from WALS (Evans 2009)

On a world scale, the countries containing the most linguistic diversity, and hence posing the greatest challenges for documentary linguistics in the next decades, are largely those with severe limits on this type of infrastructure. Whether diversity is measured by sheer number of languages, or by deep-time lineages, the result is the same:

- only three of the 25 most linguistically diverse countries in the world are OECD countries (Australia, USA and Canada).¹
- only eight of the 25 most linguistically diverse countries in the world are G20 countries (adding Brazil, China, India, Indonesia, Mexico, Russia to the above three)

On the other hand, a number of other countries with strong descriptive traditions over the last two decades are not themselves hotbeds of indigenous linguistic diversity:

Germany, France, Japan, Netherlands, UK

3. The scale of the challenge in a little more detail: Papuan languages

At least 800 languages and 80 clades, breaking into Trans-New Guinea (the world's 3rd largest language family by number of languages and non-TNG, which cumulatively represent the greatest lineage diversity on earth)

3.1 Non-TNG

Approx **30** non-Trans New Guinea families / isolates (likely to be many more as better data comes in)

Of these, only about **3** have at least one high-quality published grammar (Yimas, Manambu, Lavukaleve) and a slightly larger number have a high-quality unpublished description (Skou, Motuna, Yeli-Dnye, Meryam)

3.2 TNG

Within the Trans-New Guinea phylum, around 50 major clades (at levels comparable to Germanic)

Of these, around 15 are represented by at least one decent grammar

E.g. Usan (Reesink), Hua (Haiman), Mian (Fedden), Oksapmin (Loughnane), Dani (Bromley), Amele (Roberts), Kobon (Davies), Fore (Scott)

Overall:

- c. 800 Papuan languages
 - < 20 good grammars
- c. 80 mid-level genetic groupings within Papuan
 - < 20 represented by even one good grammar

3.3 Zooming in on one particular part of New Guinea: Southern New Guinea

¹ See Harmon (1996) for a ranking by numbers of languages, and Evans (2009) for a ranking by lineages.

Unpacking Morehead-Wasur

Morehead-Wasur

c. 20 languages in 4 main branches (probably): Kanum, Yei, Tonda, Nambu

Minor sketch material on some (e.g. Boevés' sketch grammar of Arammba [Tonda], articles by Donohue on some aspects of Kanum)

For most there is essentially no material

Pahoturi Rivers:

c. 4 languages, closely related

Small amounts of unverified sketch material, e.g. orthographic sketches or small amounts of unanalysed material by SIL linguists

4. Learning from the last 40 years: a survey of high-quality grammars

Following remarks are based on my personal 'top 50' – grammars I know reasonably well, judge highly, and refer to regularly in typological work. (I have excluded grammars of well-known languages like Japanese, English, Tagalog, Hausa and even Samoan because the focus here is on getting descriptions of underdescribed languages)

Notable about this list:

4.1 career level

Top grammars get written by researchers at all career levels:

(a) biggest single category are grammars initially written as PhD theses (and then often published later with some revisions; * indicates cases where there was substantial subsequent work before publication), e.g. (from a rather large sample)

Alaaba (Schneider-Blum; Köln)

Ambae (Hyslop; ANU)

Benchnon (Rapold; Leiden)

Diyari (Austin; ANU)

*Dolakha Newar (Genetti; UCSB)

Gooniyandi (McGregor, Sydney)

Hup (Epps, Virginia / MPI-EVA Leipzig)

Iraqw (Mous, Leiden)

Jahai (Burenhult; MPI Nijmegen)

Jamul Tipay (Miller; UCSD)
 *Kamaiura (Seki; Campinas)
 Kayardild (Evans; ANU)
 Kwaza (van der Voort; Leiden)
 Lavukaleve (Terrill; ANU)
 Manam (Lichtenberk, Hawaii)
 Mongsen Ao (Coupe, LaTrobe)
 Motuna (Onishi ; ANU)
 Mian (Fedden; Melbourne)
 Movima (Haude; Leiden)
 Mparntwe Arrernte (Wilkins, ANU)
 Mundari (Osada, Ranchi)
 Mwotlap (Francois; Paris)
 Ngiyambaa (Donaldson, ANU)
 Savosavo (Wegener; MPI Nijmegen / Radboud)
 Semelai (Kruspe; Melbourne)
 *Slave (Rice; Toronto)
 South Efate (Thieberger, Melbourne)
 Tinrin (Osumi; ANU)
 Tirio (Meira; Leiden)
 Tuscarora (Mithun; Yale)
 Udihe (Nikolaeva)
 Ungarinyin (Rumsey; Chicago)
 Wolaitta (Azeb Amha; Leiden)

[Note also that this list only includes works published from the mid-60s on. If the previous couple of decades were included, the national balance would shift substantially, reflecting the large number of doctoral grammars published in the US during the structuralist era, e.g. Sapir's grammar of Southern Paiute, right up to e.g. Rabel's grammar of Khasi (early 1960s; Louisiana). The rise of generativism dealt a huge blow to doctoral grammar-writing inside the US [with a few exceptions like Chicago, UCSB, UCSD and Oregon] and other countries emulating its policies)

(b) PhD Descriptions based on missionary work

The above examples were people who normally started fresh on the language at the inception of doctoral study. However, there is another significant category: missionaries who had been learning the language for many years but used the opportunity of a PhD to convert this into a grammar, e.g.

Anga (Whitehead; Manitoba)
 Hixkaryana (Derbyshire; ?)
 Kobon (Davies; ?)
 Usan (Reesink; Leiden)

However, there are many other categories:

(c) **PhD spinoffs** (i.e. researcher did a different topic for their PhD, but learned a language in the course of other research and later wrote a grammar):

Belhare (Biekel; MPI Nijmegen)
Lao (Enfield; ANU/Melbourne / Nijmegen)
Lezgian (Haspelmath; Freie Universität Berlin)
Nunggubuyu (Heath; Chicago)
Nishnaabemwin (Valentine; ?)

(d) **Postdocs** (fewer than I expected)

Andoke (Landaburu, CNRS Paris)
Nelemwa (Bril, CNRS Paris)
Ngalakan, Mangarayi (Merlan; AIATSIS)
Ngandi, Mara, Warndarang (Heath, AIATSIS)

(e) **Teaching academics, early to mid career**

Bininj Gun-wok (Evans; Melbourne)
Hua (Haiman; ANU)
Qiang (LaPolla; LaTrobe)
Yidiny (Dixon; ANU)
Yimas (Foley; ANU)

(f) **Teaching academics, mid-to-late career (or lifetime cumulative)**

Cupeño (Hill; Tucson)
Jamsay (Heath; Michigan)
Musqueam (Suttles; UBC)
Toqabaqita (Lichtenberk; Auckland)

(g) **Research academics, mid-to-late career**

Jarawara (Dixon; LaTrobe RCLT)
Manambu (Aikhenvald; LaTrobe RCLT)
Hdi, Lele etc (Frajzyngier; Boulder)

(h) **Research teams (ranging from seasoned researchers to undergrads)**

Tsakhur Kibrik et al
Bagvalal Kibrik et al

4.2 School and descriptive tradition

Several schools / traditions have had a disproportionate influence on the field:

Boasian / Sapirian structuralism (USA, 1920s-early 1960s)

ANU Linguistic Traditional (ANU, Melbourne, Latrobe, 1970s-present)

Dutch linguistic tradition (Leiden, 1960s - present)

Russian / Soviet descriptive tradition (Moscow and St Petersburg; 1930s - present)

Each of these has

- placed grammar-writing at the heart of its conception of linguists should be doing
- encouraged the writing of grammars as PhD dissertations

4.3 Type of project / funding environment

Significant clusters of grammars in the above list reflect large-scale projects aimed at producing grammars of little-documented regions, e.g.

- Dutch project (led by Muysken) on languages of Bolivia etc. (e.g. Haude)
- French project (CNRS; led by Jon Landaburu) on languages of Colombia
- long-running project on Australian languages (Dixon at ANU, 1970s and 1980s plus AIATSIS postdocs in early 1970s)
- RCLT project on Amazonian languages (RCLT, LaTrobe)
- German project on Cushitic grammars (Cologne)

4.4 Publishers

Only a small number of publishers produce high-quality grammars of little-described languages in large numbers

Mouton de Gruyter (Berlin)

Pacific Linguistics (Canberra)

UC Press (Berkeley)

U Hawaii Press (Hawaii)

Helmut Buske Verlag (Hamburg)

Several others have had short-lived grammar series which were then discontinued owing to perceived financial non-viability, e.g. Cambridge UP (discontinued twice, first in the Blue Series (4 classic Australian grammars between 1972-1981, then nothing); then in the 2000s (Cambridge Grammar Series, which published several magnificent grammars, e.g. Kruspe's grammar of Semelai and Watters' grammar of Kham)

A number of other publishers have experimented with one or two good grammars (e.g. Stanford U. Press for Yimas; OUP for Jarawara) without establishing a regular series

4.5. Caveat: alternatives to grammars: analysis clusters

Warlpiri

[The best-known and best-understood Australian language despite the absence of a reference grammar. Scholars have to integrate large numbers of papers; glosses, analyses and terms vary from paper to paper, but there has been detailed exploration of topics at levels that far exceed what we have for any other Australian language. (Including at least 5 PhDs on aspects of Warlpiri).

A huge and exemplary dictionary of Warlpiri is still in preparation

Kalam

Early grammar by Pawley (1966), fairly inaccessible; most of the running on Kalam is found in a string of insightful and exploratory papers, supplemented by one MA (Lane; Auckland) and a nearly-finished dictionary

[Pitting Warlpiri against Pitjantjara, or Kalam against Kobon, the information is far more detailed and insightful for the first member of each pair despite the existence of full grammars for the second, though more difficult to decide on an integrated analysis]

Jaminjung

PhD by Schultze-Berndt has a sketch grammar in one chapter before passing to detailed explanation of one topic (coverb constructions); other materials on J. can be found in various papers and there is now a large DoBeS style corpus

Other examples might be *Tzotzil* (Aissen book on grammatical relations)

6. Some personal observations on what supports grammar-writing

- Academic cultures (typically national) placing a high value on grammar-writing at all levels
- Cohort
- Culture of grammar-reading
- Close study of one or more exotic languages
- study of closely related languages (and comparative analysis) of the German tradition of a Strukturkurs
- accumulation of tests / argumentation relevant to phenomena in particular language types/families
- targeted projects (cf the Muysken project on languages of Bolivia)

- centres supporting grammar write-up (cf RCLT at LaTrobe)
- time out for analysis (one year write-up support), particularly important for academics later in their careers, but also to support PhD by-products and to help deal with the ‘truncated postdoc syndrome’)
- funding support for grammar projects through national and international agencies
- even-handed ideology that values both outsider contributions (currently typical) and insider ones (the hoped-for contribution of trained native speakers, which still remains a largely unfulfilled ideal)

7. Has the documentarian movement helped grammar writing?

- theoretical / ideological hindrances

Himmelman (2006):

linguists should forego the writing and publication of descriptive grammars in order to devote their full time and energies to the task of language documentation:

The writing of a descriptive grammar involves to a substantial degree matters of formulation ... and organization... These are very time consuming activities which in some instances may enhance the analysis of the language system, but in general do not contribute essential new information on it. Thus, with regard to the economy of research resources, it may be more productive to spend more time on expanding the corpus of primary data rather than to use it for writing a descriptive grammar (Himmelman 2006:24)

Mosel (2006:307) on sketch grammars:

‘a thorough analysis can wait until there is time for a specialist investigation’. This ‘let the grammar wait’ position is based on a claim that

Cf Rhodes et al (2006):

Himmelman (1998) has argued persuasively that documentation is distinct from what he calls description, i.e. linguistic analysis. We think this is seriously mistaken. In order to know how far along one has come in documenting a language, one must be able to measure how far there is to go. A crucial part of that measurement is found in the accounting function of analysis. How do we know when we’ve gotten all the phonology? When we’ve done the phonological analysis and our non-directed elicitation isn’t producing any new phonology.

- practical hindrances

time on metadata, coding, transcription etc. vs. time on analysis

quantity-based deposit requirements

shift to more realistic but messy data makes it harder to formulate clear generalisations

- training
is the time spent on training documentary techniques detracting from that spent on learning analysis, studying how closely-related languages work, etc.?

8. The grammar-writing challenge: a seven-point manifesto

(i) recognition of scale and urgency: a fivefold increase

We need a five- to ten-fold upscaling of our efforts over the next 2-3 decades if we are to come anywhere near the challenge of getting good grammatical descriptions of most of the world's languages while we can

(ii) international goal-keeping

With the Human Genome project we knew how far we had to go –
We need equivalent, widely publicised goals at world-wide level in order to monitor and direct our progress

(iii) placing grammar-writing as a central goal of linguistics

- post-UG paradigm change is happening that restores the study of linguistic diversity to the centre of the field's goals
- paradoxically, the surge of support for language documentation in the last decade has not really helped grammar-writing; this also needs to be turned around through a retheorization of the language documentation process, away from documentarist fundamentalism
- the last few decades makes it clear that good grammar-writing very directly reflects overall valuation of the activity at the level of national scholarly communities (with an additional kick from international recognition)
- funding priorities in many countries and international organisations need to revalue grammar-writing and the research that leads to it (positive examples are recent initiatives in the NSF, and the BABEL initiative of the ESF)
- prizes etc. can help (e.g. the Gabelentz and Panini awards by the ALT, use of awards like the Bloomfield award for grammars) but overall professional valuing is much more important

(iv) broad-spectrum support for different career categories

- to get anywhere near our goals, we need to support grammar-writing at all career stages. What is needed varies according to the level (PhD candidates have time but need mentoring and input; for later-career researchers the needs are reversed. Likewise, field logistics are more likely to involve two or three lengthy periods for PhD students, but a larger number of shorter periods for later-career researchers)

Steps are obvious:

- more funding for PhD grammar projects
- more targeted postdocs, with realistic time frames (the standard 2-year postdoc is too short)
- more support for fieldwork time-release for mid-career researchers
- more support for one-year write-up fellowships on the RCLT model (all 2009 Gabelentz finalists received RCLT support)

(v) teaching and training

- recognition needed that teaching Field Methods and Documentary Techniques does not of itself teach grammar writing – both are needed, but the goals are distinct
- we don't really know what types of teaching produce optimal grammar-writing (attempts to teach at as a separate subject are in their infancy and in any case may not be the best way).

However, it's pretty clear that it will include

- specific problem-solving (the classic linguistics data problem)
- typological training (especially in 'General Comparative Grammar) to widen conceptions of what to look for
- integrative in-depth study of one or two exotic languages as the best way of teaching how synthetic descriptions are produced: *grammar-reading and grammar-study precedes grammar-writing!*

(vi) international collaboration to multiply the number of centres giving good training

- need for expansion in training programs at all levels (Bachelors through to PhD) that will recruit and train linguists from high-linguistic-diversity countries
- internationally-coordinated summer-schools and other on-site intensive programs hold immense potential – is it possible to develop an international multi-university initiative that hooks these up to degree-level training through a mix of on-site and external training?
- two interesting models – the African Institute of Mathematics in Capetown and OKMA as a national-level initiative in Guatemala (now unfortunately phased out)

- DoBeS and Rausing programs show the power of well-funded international initiatives; we need more of these (and oriented to grammar-writing)

(vii) publishing support

We need to find ways of persuading more publishers to take on grammars

At the same time we need to recognise that they are businesses and can't be expected simply to run at a loss (MdeG cross-subsidises from other publication lines): the linguistics profession needs to

- drum up regular publishing subsidy schemes
- make sure more linguistics libraries (and more linguists!) subscribe to grammar series

References

- Evans Nicholas. 1995. *A Grammar of Kayardild*. Berlin: Mouton de Gruyter..
- Evans, Nicholas. 2008. Review of Gippert, Jost, Nikolaus Himmelmann and Ulrike Mosel (eds.), *Essentials of language documentation*. Berlin / New York: Walter de Gruyter. *Language Documentation and Conservation* 2.2: 340-350.
- Evans, Nicholas. 2009. *Dying Words: Endangered languages and what they have to tell us*. Maldon & Oxford: Wiley-Blackwell. The Language Library. pp. xx + 267.
- Gippert, Jost, Nikolaus Himmelmann and Ulrike Mosel (eds.), 2007. *Essentials of language documentation*. Berlin / New York: Walter de Gruyter.
- Himmelmann, Nikolaus. 2007. Language documentation: what is it and what is it good for? In Gippert et al, pp. 1-30.
- Mosel, Ulrike. 2007. Sketch grammars. In Gippert et al, p. 301-310.
- Mithun, Marianne. 2001. Who shapes the record: the speaker and the linguist. In Paul Newman and Martha Ratliff (eds.), *Linguistic Fieldwork*. Cambridge: CUP, pp.34-54.
- Morey, Stephen. 2005. *The Tai languages of Assam – a grammar and texts*. Canberra: Pacific Linguistics.
- Thieberger, Nicholas. 2006. *A grammar of South Efate*. Honolulu: University of Hawaii Press.
- Zaefferer, Dietmar. 2006. Realising Humboldt's dream: cross-linguistic grammatography as database creation. In Felix Ameka, Alan Dench and Nicholas Evans (eds.) *Catching Language: the standing challenge of grammar-writing*. Berlin: Mouton de Gruyter.