

Diana Santos · Krister Lindén
Wanjiku Ng'ang'a *Editors*

Shall We Play the Festschrift Game?

Essays on the Occasion of Lauri Carlson's
60th Birthday

 Springer

Shall We Play the Festschrift Game?

Diana Santos • Krister Lindén • Wanjiku Ng'ang'a
Editors

Shall We Play the Festschrift Game?

Essays on the Occasion of Lauri Carlson's
60th Birthday

 Springer

Editors

Diana Santos
Department of Literature, Area Studies
and European Languages
University of Oslo
Oslo, Norway

Wanjiku Ng'ang'a
School of Computing & Informatics
University of Nairobi
Nairobi, Kenya

Krister Lindén
Department of Modern Languages
University of Helsinki
Helsinki, Finland

ISBN 978-3-642-30772-0

ISBN 978-3-642-30773-7 (eBook)

DOI 10.1007/978-3-642-30773-7

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012946035

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Playing the Festschrift Game

1 Presentation

There are not many people who can be said to have influenced and impressed researchers in so many disparate areas and language-geographic fields as Lauri Carlson, as is evidenced in the present Festschrift.

Lauri's insight and acute linguistic sensitivity and linguistic rationality have spawned findings and research work in many areas, from non-standard etymology to hard-core formal linguistics, not forgetting computational areas such as parsing, terminological databases, and, last but not least, machine translation.

In addition to his renowned and widely acknowledged insights on tense and aspect, and their relationship with nominal quantification, and his ground-breaking work in dialogue using game-theoretic machinery, Lauri has in the last fifteen years as Professor of Language Theory and Translation Technology contributed immensely to areas such as translation, terminology and general applications of computational linguistics.

The three editors of the present volume have successfully performed doctoral studies under Lauri's supervision, and wish with this volume to pay tribute to his supervision and to his influence in matters associated with research, and scientific, linguistic and philosophical inquiry, as well as to his humanity and friendship as a definitely non-bossy supervisor.

Diana still recalls with gratitude the short periods either in Helsinki or in Lisbon, where discussions on tense and aspect matters took place:

The pertinence of Lauri's questions, questioning what at the time seemed unquestionable matters to me. The need to try to explain seemingly evident things in my understanding of my own native language was a sobering and extremely useful exercise. His genuine interest in language in general, and languages in particular, was also contagious, and the turning point in my PhD study—and possibly my whole research career—came about by Lauri buying a cheap edition of Steinbeck's *The Pearl* and its translation into Portuguese, and challenging me to prove that my translation rules were backed by real-world evidence. Naturally, as soon as we started looking at the texts themselves, a whole new world opened. Parallel corpora studies can now be said to constitute a well-established and productive field. Lauri does not even claim to have played a serious part in it. But he will clearly remain for me one of its founders, or rather one of its most important sources of inspiration.

After his graduation, Krister was involved in an English-to-Finnish Machine Translation project under the supervision of Lauri Carlson, who later also became the primary supervisor of his thesis:

His vast knowledge of both applied and formal linguistics combined with his down-to-earth remarks paved my way to linguistics. Later, Lauri again proved his worth in gold when he, with a few keystrokes, fleshed out a definition of synonymy and how to implement it in practice for our joint project on creating a Finnish wordnet, which is now the second largest wordnet in the world after the famous Princeton WordNet.

Given her non-linguistic background at the start of her PhD study, Wanjiku is hugely appreciative of Lauri's mentorship and patience that saw her acquire the requisite linguistic knowledge, that enabled her to work with Swahili:

Working with a then resource-scarce language required clever combination of resources and methods, and Lauri was very supportive of my lines of inquiry, a testament to his liberal thinking and open-mindedness, and I remember our meetings fondly. Our varied discussions on language, culture and religion helped me understand and appreciate the intricacies of language and the strengths and limitations of machine learning approaches to language processing. In addition, I benefited immensely from exposure to some of his networks—collaborative linkages which were crucial for my research, and others which persisted after my doctoral studies and which continue to shape my career to date.

2 Book Contents

Here, we provide a brief overview of the book's contents by describing, for each paper, the pertinent research questions and the points of contact with the other texts. This brief overview serves to harmonise the various contributions, and should in no way diminish the need to read each chapter individually.

The general nature of linguistic inquiry and how methods and empirical data should be tackled is the subject of Fred Karlsson's contribution, *Is There a Crisis in Generative Linguistics?*, an answer to a paper by Lauri on a similar subject. A quite original and interesting related subject is the way etymology is construed in different traditions, with David Nash's thought-provoking piece on language history in Australia, *"It's Etymology Captain, But Not As We Know It": 'Pump' in North Australia*.

Translation in History by Nicholas Ostler takes us on the challenging endeavor of making sense of translation diachronically and provides a new dimension to an ever-present activity in our times. Andrew Chesterman's chapter, *Catford Revisited*, provides a new view on Catford, an early theorist of translation who deserves to be rediscovered and reappraised, in his view. Diana Santos's contribution, *The Next Step for the Translation Network* delineates a computational program for studying human translations based on large corpora, as a direct follower to the proposal sketched in her dissertation.

An intriguing investigation on the concept of core vocabulary is provided by Lars Borin's *Core Vocabulary: A Useful But Mystical Concept in Some Kinds of Linguistics*, which provides thought-provoking material on some basic concepts in need

of rethinking. On a practical note, Krister Lindén, Jyrki Niemi and Mirka Hyvärinen provide an interesting evaluation of the Finnish wordnet and its coverage of the Finnish core vocabulary in *Extending and Updating the Finnish Wordnet*. As mentioned, Lauri Carlson was instrumental when creating the initial version of the Finnish wordnet. An exciting related subject, in a different frame of mind, is Janet Pierrehumbert's work on *Burstiness of Verbs and Derived Nouns*, which merges formal semantics and statistical insights to Lauri's probably dearest subject of all, discourse.

Creating cutting-edge linguistic-computational resources, both for Finnish and for other languages, is the subject of *Outsourcing Parsebanking: The FinnTreeBank Project* by Atro Voutilainen, Tanja Purtonen and Kristiina Muhonen, where the authors suggest and demonstrate a novel way for developing large treebanks. In order to assist the treebanking effort, Anssi Yli-Jyrä's article *On Dependency Analysis via Contractions and Weighted FSTs* deals with an improved parsing technique for dependency parsers, presenting a system which is a radical improvement on his dissertation work, of which Lauri was one of the supervisors.

New analyses of linguistic phenomena round up this widely varied book: Patrick McConvell and Jane Simpson discuss some intriguing features of case marking in *Fictive Motion Down Under: The Locative-Allative Case Alternation in Some Australian Indigenous Languages*, describing several forms to mark the different location of events and of participants, and proposing a grammaticalisation path, while also comparing with parallel phenomena in Finnic. Aet Lees studies the expression of obligation and need in this last language family using Bible translations, in her chapter *Necessive Expressions in Finnic Bible Translations*, which handles the subject diachronically.

Wanjiku N'gang'a's *Building Swahili Resource Grammars for the Grammatical Framework*, in addition to the theoretical part of resource grammar writing, reports on an aid for visually-impaired mobile phone users that has been developed using the Grammatical Framework. Her work builds on the Grammatical Framework of Aarne Ranta, whose chapter *On the Syntax and Translation of Finnish Discourse Clitics* not only discusses Finnish discourse clitics but also describes a Finnish-English machine translation system that handles this interesting issue in Finnish grammar. In fact, it should be mentioned that this last chapter achieves the highest number of connections to Lauri's main interests: discourse, translation, formal analysis, and computation.

3 Contributions

In addition to the authors, a number of people have also contributed to this book in various capacities. Maria Vilkuna and Antti Arppe helped in reviewing the contributions. Jyrki Niemi undertook the job of converting all the non- \LaTeX contributions to \LaTeX and unifying the ones provided in \LaTeX . Finally, we are grateful to Seija Carlson for her help and for keeping it all a secret!

We would also like to acknowledge Lauri's friends and colleagues who would have wanted to write a chapter but did not manage to do so—either because we were unable to contact them and they consequently were unaware of this noble initiative, or due to pressing time constraints.

We are also grateful to all the authors, who willingly participated in the cross-review process, providing valuable feedback which has undoubtedly resulted in the high quality and readability of this book. We as editors are therefore extremely grateful for this added value.

Acknowledgements Diana Santos acknowledges the support of the University of Oslo in preparing this volume, and Krister Lindén would like to thank the FIN-CLARIN infrastructure project at the University of Helsinki.

Oslo
Helsinki
Nairobi

Diana Santos
Krister Lindén
Wanjiku Ng'ang'a

Contents

Is There a Crisis in Generative Linguistics?	1
Fred Karlsson	
“It’s Etymology Captain, but Not as We Know It”: <i>Pump</i> in North Australia	7
David Nash	
Translation in History	19
Nicholas Ostler	
Catford Revisited	25
Andrew Chesterman	
The Next Step for the Translation Network	35
Diana Santos	
Core Vocabulary: A Useful But Mystical Concept in Some Kinds of Linguistics	53
Lars Borin	
Extending and Updating the Finnish Wordnet	67
Kristen Lindén, Jyrki Niemi, and Mirka Hyvärinen	
Burstiness of Verbs and Derived Nouns	99
Janet B. Pierrehumbert	
Outsourcing Parsebanking: The FinnTreeBank Project	117
Atro Voutilainen, Tanja Purtonen, and Kristiina Muhonen	
On Dependency Analysis via Contractions and Weighted FSTs	133
Anssi Yli-Jyrä	
Fictive Motion Down Under: The Locative-Allative Case Alternation in Some Australian Indigenous Languages	159
Patrick McConvell and Jane Simpson	

Necessive Expressions in Finnic Bible Translations 181
Aet Lees

Building Swahili Resource Grammars for the Grammatical Framework . 215
Wanjiku Ng’ang’a

On the Syntax and Translation of Finnish Discourse Clitics 227
Aarne Ranta

Contributors

Lars Borin Språkbanken, Department of Swedish, University of Gothenburg, Gothenburg, Sweden

Andrew Chesterman Translation Studies, Department of Modern Languages, University of Helsinki, Helsinki, Finland

Mirka Hyvärinen Department of Modern Languages, University of Helsinki, Helsinki, Finland

Fred Karlsson General Linguistics, Department of Modern Languages, University of Helsinki, Helsinki, Finland

Aet Lees Department of Linguistics, University of Sydney, Sydney, Australia

Krister Lindén Department of Modern Languages, University of Helsinki, Helsinki, Finland

Patrick McConvell School of Language Studies, Australian National University, Canberra, Australia

Kristiina Muhonen Department of Modern Languages, University of Helsinki, Helsinki, Finland

David Nash School of Language Studies, The Australian National University, Canberra, Australia

Wanjiku Ng'ang'a School of Computing and Informatics, University of Nairobi, Nairobi, Kenya

Jyrki Niemi Department of Modern Languages, University of Helsinki, Helsinki, Finland

Nicholas Ostler Foundation for Endangered Languages, Bath, England, UK; School of Oriental and African Studies, London, England, UK

Janet B. Pierrehumbert Department of Linguistics, and Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA

Tanja Purtonen Department of Modern Languages, University of Helsinki, Helsinki, Finland

Aarne Ranta University of Gothenburg, Gothenburg, Sweden

Diana Santos University of Oslo, Oslo, Norway

Jane Simpson School of Language Studies, Australian National University, Canberra, Australia

Atro Voutilainen Department of Modern Languages, University of Helsinki, Helsinki, Finland

Anssi Yli-Jyrä Department of Modern Languages, University of Helsinki, Helsinki, Finland

Is There a Crisis in Generative Linguistics?

Fred Karlsson

Abstract Several recent critiques have claimed that the empirical foundations of generative linguistics are weak due to the reliance on informally gathered grammaticality judgments drawn from the intuitions of the researcher. Phillips (In: Japanese/Korean linguistics, vol. 17, 2009) argued i.a. that, in order for there to be a theoretical crisis, two criteria should be fulfilled, namely (i) intuitive judgments have led to generalizations that are widely accepted yet bogus, and (ii) misleading judgments form the basis of important theoretical claims or debates. Furthermore Phillips claimed that (i, ii) have not been fulfilled. I argue that (i, ii) have in fact been satisfied because Chomsky's (Q. Prog. Rep.—Mass. Inst. Technol., Res. Lab. Electron. 41:64–65, 1956; Syntactic structures. Mouton, The Hague, 1957) intuition-based claim that English is not a finite-state language is demonstrably false.

In his article “Should we impeach armchair linguists?”, Phillips (2009) asked whether there is a crisis in generative linguistics. Several recent critiques had claimed that the empirical foundations of generative linguistics are weak due to the reliance on informally gathered grammaticality judgments drawn from the intuitions of the researcher. Here is Phillips' answer:

The claim is not just that one finds questionable examples in linguistics papers, but that lax data-collection standards have led to the growth of theories that are based upon bogus data. If these charges stick, then we face a genuine crisis. In order for there to be a crisis, however, it would need to be the case that (i) Intuitive judgments have led to generalizations that are widely accepted yet bogus. (ii) Misleading judgments form the basis of important theoretical claims or debates. (iii) Carefully controlled judgment studies would solve these problems. Although I sympathize with the complaint that one can find many cases of poor data collection in the linguistics literature, I am not sure that any of (i)–(iii) is correct. A surprising number of the critiques that I have read present no evidence of the supposed damage that informal intuitions have caused, and among those that do provide specific examples it is rare to find clear evidence of the lasting impact of questionable judgments. (Phillips 2009: 3)

Phillips thus held that criteria (i, ii) have not been satisfied. As for (iii), he argued that experimental data from large-scale rating studies and acceptability tests

F. Karlsson (✉)

General Linguistics, Department of Modern Languages, University of Helsinki, Helsinki, Finland
e-mail: fred.karlsson@helsinki.fi

“are likely to be less of a panacea than we are sometimes led to believe” even if he acknowledged the value of experimentation as a supplementary data-gathering method, alongside introspection.

My thesis here is that (i) and (ii) are in fact satisfied. My test-case will be one of the fundamental claims of generative grammar, articulated by Chomsky (1956: 65), that English is not a finite-state language, plus one of its corollaries. Here is Chomsky’s whole argument which I shall call *English-not-FSL*:

Turning now to the English language, we find that there are infinite sets of sentences with just the mirror-image properties of [the artificial language] L_1 . For example, let S_1, S_2, S_3, \dots , be declarative sentences. Then the following are all English sentences:

- (1) (i) If S_1 , then S_2 .
- (ii) Either S_3 , or S_4 .
- (iii) The man who said that S_5 , is arriving today.

These sentences have dependencies between “if” and “then”, “either” and “or”, “man” and “is”. But we can choose S_1, S_3 , and S_5 in (1) as (1i), (1ii), or (1iii) themselves. Proceeding to construct sentences in this way, we arrive at sentences with dependency sets of more than any fixed number of terms, just as in the case of L_1 . English is therefore not a finite-state language. (Chomsky 1956: 65)

From this an important corollary can be deduced: syntactic recursion, especially nested recursion (center-embedding) is unlimited. This I call *Unlimited-center-embedding*. Both English-not-FSL and Unlimited-center-embedding have played an enormous role in the theoretical genesis of generative grammar. Note that the truth of Unlimited-center-embedding is the one and only criterion deciding whether natural language syntax is finite-state or context-free, i.e. whether English-not-FSL is true. If Unlimited-center-embedding is false, English-not-FSL too is false and natural language syntax is (in principle) of finite-state character.

It would be easy to list hundreds of scholarly articles, text books and overviews from linguistics and neighboring disciplines where English-not-FSL and Unlimited-center-embedding are taken to be truths established by the early Chomsky. Here are some representative examples from theoretical, mathematical, computational and psycholinguistics:

First we will observe that English has characteristics that put it outside the class of regular languages. Then we will consider a simple theory of syntax that is essentially like the theory of finite automata or finite state grammars, and conclude that the theory is inadequate (Chomsky 1956, 1957). (Bach 1974: 187)

The demonstration that English is not a finite automaton language was one of the first results to be achieved in the nascent field of mathematical linguistics (Chomsky 1956, 1957, Chap. 3). . . . This illustrates one sort of practical result that can sometimes be obtained from the study of formal languages. (Partee et al. 1993: 478–479)

Chomsky (1957) showed that natural language cannot be characterised by a finite-state device. In particular, it cannot produce arbitrarily long sequences of multiple center-embedded structures. (Harley 1995: 152)

Every computational linguistics graduate student knows, from the first textbook that introduces the Chomsky hierarchy, that English has constructs, such as center embedding, that cannot be described by any finite-state grammar. (Hobbs et al. 1997: 383)

The fact that a sentence with, say, ten dependent clauses is unlikely to be spoken and would be impossible to understand at first hearing does not detract from Chomsky's argument that a recursive (i.e. infinitely repeatable) rule of embedding must, in principle, be allowed for in English grammar. Clearly, it would be arbitrary to say that embeddings can only be carried out three times or that sentences can only be of a certain length. (Greene 1972: 26)

Chomsky obviously intuited his examples (1i–iii) as well as his subsequent claims about sentences of unlimited embedding complexity; at least he did not report that his analysis was based either on observation of language use (i.e. corpus work) or on experimentation. Surely English-not-FSL and Unlimited-center-embedding are widely accepted generalizations in the sense of Phillips' criterion (i), and also important theoretical claims as meant by his criterion (ii). Many (all?) generative linguists still subscribe to Unlimited-center-embedding and English-not-FSL in the spirit just quoted, despite counterarguments of critics like Levelt (1974) and Manaster Ramer (1983). For example, the hypothesis by Hauser et al. (2002), that the faculty of language in the narrow sense (FLN) only includes recursion, is an extended variant of Unlimited-center-embedding.

The crucial question is whether Unlimited-center-embedding is true. Using the third possible data-gathering method, extensive corpus study of i.a. English, Finnish, German, Latin and Swedish, I demonstrated in Karlsson (2007), first (a) that multiple center-embedding of clauses is practically non-existing in spoken language; (b) that the limit of multiple center-embedding in written language is three, but such constructions are extremely rare and mostly more than 100 years old, making double center-embedding the only reasonably productive type of center-embedding; and (c) of the extant double center-embeddings (more than 100 authentic instances), contrary to the hypothesis of Miller and Chomsky (1963), the most common type of multiple center-embedding is *self-embedding* with two relative clauses; this is the only type of central self-embedding that occurs.

My interpretation of these empirical data is that they falsify the hypothesis of Unlimited-center-embedding and considerably downplay the general significance of recursion, especially its theoretically most important manifestation as nesting (i.e. center-embedding); of course, left- and right-branching recursion is equal to iteration. Consequently, English-not-FSL too is false, English (and other languages) are regular languages. Upholding the idea of Unlimited-center-embedding at competence level, and relegating the restrictions on center-embedding to the domain of performance, begs the question and makes the hypothesis of Unlimited-center-embedding immune to criticism.

Finally, let us reiterate Phillips' two first criteria:

In order for there to be a crisis, however, it would need to be the case that (i) Intuitive judgments have led to generalizations that are widely accepted yet bogus. (ii) Misleading judgments form the basis of important theoretical claims or debates.

Both (i) and (ii) are demonstrably at hand—and, ergo, so is the crisis. Of course, it is quite another matter whether the community of practitioners of generative linguistics is willing to admit this state of affairs. When I submitted the above comments for consideration to be published in the journal *Linguistic Inquiry* in June 2010, the editors needed only three days to refuse my offering with the main argument that there

was nothing new in my analysis because Karlsson (2007) already had presented the relevant empirical data. I consider this a classical instance of gate-keeping: research paradigms (in their Kuhnian ‘senile phase’ of development) are not to be disturbed by outsiders.

Now that I have the privilege of publishing these remarks in Lauri Carlson’s *Festschrift*, I want, by way of conclusion, to briefly address an issue he brought up in his contribution (Carlson 2006) to my *Festschrift*. Lauri launched the highly interesting idea that spoken language syntax is closer to written language syntax than many linguists claim these days, namely given that spoken language is analyzed with due consideration being paid to the immediate discourse context (which e.g. helps to fill in ellipses). This idea I find intriguing and hopefully it will be fleshed out some day.

There is just one detail I want to take issue with. Lauri claimed (Carlson 2006: 128) that one “... expect[s] to find in real speech artful periods like” (here cited in the original form from Albert Engström’s journal *Strix*, 1902):

Om dig, Eriksson, om hvilken jag ej kunnat tänka mig något dylikt, har jag, då du, då jag till följd av iråkad snuva, nös, skrottade, ändrat uppfattning.

‘Of you, Eriksson, whom I would not have thought capable of it, I have, now that you, when I had to sneeze due to a head cold, burst into a laugh, changed my opinion.’

My rejoinder here is that such a sentence is, in fact, extremely unlikely to occur even in writing because it displays a type of multiple center-embedding that has (so far) not been attested in any type of authentic language use, that is self-embedding of two adverbial clauses of the same type (two *då* ‘when’ clauses). The primary purpose of Engström’s journal *Strix* was to excel in enlightened humor. So here we are not dealing with ordinary spoken language but rather with a deliberate norm breach of a soft constraint in written language (the ban on certain types of self-embedding) and it is this breach which generates the humorous effect.

References

- Bach, Emmon. 1974. *Syntactic theory*. New York: Holt, Rinehart and Winston.
- Carlson, Lauri. 2006. Parsing spoken dialogue. In *A man of measure: Festschrift in honour of Fred Karlsson. on his 60th birthday*, eds. Mickael Suominen, Antti Arppe, Anu Airola, Orvokki Heinämäki, Matti Miestamo, Urho Määttä, Jussi Niemi, Kari K. Pitkänen, and Kaius Sinnemäki. Vol. 19 of *SKY journal of linguistics*, 128–137. Turku: The Linguistic Association of Finland. <http://www.linguistics.fi/julkaisut/sky2006special.shtml> (special supplement).
- Chomsky, Noam. 1956. On the limits of finite-state description. *Quarterly Progress Report—Massachusetts Institute of Technology. Research Laboratory of Electronics* 41: 64–65.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- Greene, Judith. 1972. *Psycholinguistics*. Harmondsworth: Penguin.
- Harley, Trevor. 1995. *The psychology of language. From data to theory*. Hove: Erlbaum.
- Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 298: 1569–1579.
- Hobbs, Jerry R., Douglas Appelt, John Bear, David Israel, Megumi Kameyama, Mark Stickel, and Tyson Mabry. 1997. FASTUS: A cascaded finite-state transducer for extracting information from natural-language text. In *Finite-state language processing*, eds. Emmanuel Roche and Yves Schabes, 383–406. Cambridge: The MIT Press.

- Karlsson, Fred. 2007. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43: 365–392.
- Levelt, Willem J.M. 1974. *Formal grammars in linguistics and psycholinguistics. Vol. II: Applications in linguistic theory*. The Hague: Mouton.
- Manaster Ramer, Alexis. 1983. Soft formal underbelly of theoretical syntax. In *Papers from the 19th regional meeting*, 256–262. Chicago: Chicago Linguistic Society.
- Miller, George A., and Noam Chomsky. 1963. Finitary models of language users. In *Handbook of mathematical psychology, II*, eds. R. Duncan Luce, Robert R. Bush, and Eugene Galanter. New York: Wiley.
- Partee, Barbara H., Alice ter Meulen, and Robert E. Wall. 1993. *Mathematical methods in linguistics*. Dordrecht: Kluwer.
- Phillips, Colin. 2009. Should we impeach armchair linguists. In *Japanese/Korean linguistics*, eds. Shoichi Iwasaki, Hajime Hoji, Patricia M. Clancy, and Sung-Ock Sohn, Vol. 17. Stanford: CSLI.

“It’s Etymology Captain, but Not as We Know It”: *Pump* in North Australia

David Nash

Abstract We consider the words originating from the English word ‘bamboo’ borrowed into the indigenous languages of northern Australian and denoting the didjeridu (drone pipe) or other aerophones. The word ‘bamboo’ must have been first acquired by speakers of Australian Aboriginal languages in the 19th century, and in north Australia where the large stem plant is endemic, namely in the region of Darwin. The available data is organised in support of an hypothesised spread whereby the word was applied to the aerophone made from bamboo, and then to similar aerophones made of other wood. In this sense, ‘bamboo’ (as *pampu*) spread inland southwards, and eastwards to western Cape York Peninsula. In western Cape York Peninsula the word lost the final vowel, and in this form was borrowed southwards and applied to the particular aerophone the ‘emu caller’, used to attract the emu (a large flightless game bird). A comparable distribution is collated for an indigenous word denoting aerophones: *kurlumpu(rr)* and corresponding forms in various north Australian languages. The study demonstrates how some etymological headway can be made on loanwords in languages with only a recent documentary record.

1 Introduction

The etymology of a particular word traces combinations of inheritance and borrowing, describing shifts in form and meaning along the way. When borrowing is involved, the path can be quite idiosyncratic, and uncovering it requires a deduction

It is a pleasure to associate this paperlet with Lauri Carlson. We knew each other as fellow graduate students in linguistics at MIT, and also shared a group apartment. Lauri would occasionally put to us his roommates (mostly native speakers of English of various kinds) that a (to us implausible) pair of English words would prove to be etymologically related, and invariably on us reaching for dictionaries he would turn out to be right. An earlier version appeared as blog posts (Nash 2011a, 2011b). I am grateful to the editors for providing this opportunity, and for their forbearance.

D. Nash (✉)

School of Language Studies, The Australian National University, Canberra, Australia
e-mail: david.nash@anu.edu.au

beyond the Comparative Method. See for instance Trask (1996: 351–353) for an appraisal of the kind of study involved.

When the documentary record for the languages involved is only recent, as for Australian languages, the task is perforce more inferential. But if we confine ourselves to the historical period, that is, since written records began around two centuries ago, then etymologies in Australia can be pursued, using at least some shreds of evidence of a kind not available for the prehistoric period. In this note, I exemplify this first by studying a word of regional Aboriginal English, and then by looking at a cognate set in Australian languages with the same range of meanings. Along the way we see the hazards of translation on the colonial frontier, even with respect to ‘concrete’ artefacts.

2 ‘Pump’ in North Queensland

The published grammar of the Kalkatungu language of western Queensland has this entry in the ‘Weapons, tools, etc.’ section of the glossary:

‘pump’ (decoy device for attracting birds) **ku|umpu**¹ (Blake 1979: 179)

This entry is fairly opaque to speakers of Australian English. What kind of decoy, the reader wonders, would be named with the English word for a fluid propulsion device (let alone a kind of footwear!).

The author (BB) elaborated² that he based the entry on a sound recording of an interview with Mrs Lardie Moonlight (LM) conducted by his colleague Gavan Breen (GB), who has kindly provided me with his careful transcript of the relevant excerpt, and later played me the sound recording.

GB (<BB³): You plant so the emu won’t see you.

LM: Yes, dig a hole and put the little bough around it, sit in it with a pump, call him // *yu.ridja(y) yini / wudingalkuwa // wudingalkuwa //* he sitting down there in the hole for that emu to come; he blowing that pump, pumping his *kurumbu // kurlumbuyan i.ni*

GB: What was the pump like? How did they make it?

LM: Out of the little oller [sc. hollow] tree, they knock the hollow tree down and they put a haxe [sc. axe] round it, you know, make it small, they blow it then.

GB: Oh, yes, it makes a noise and the emu comes up to see what the noise is.

LM: Oh yeah, they make a lovely noise too.

(Breen Field Tape 283, AIATSIS AV tape A2459b, 24 May 1972, Boulia, Queensland)

The morphological analysis of the Kalkatungu is:⁴

¹ | represents l-with-dot-under, the apico-domal lateral. Blake (1979: 4) noted that ‘All the informants spoke English in most situations, some of them using a fair admixture of Pidgin features.’

² Blake and I exchanged a number of email messages on this in April–May 2011.

³ Breen was working through a set of elicitation sentences that Blake had drawn up.

⁴ Here I revert to the spellings of Blake (1979), cast in a practical orthography whereby e.g. *rl* is the apico-domal lateral; and ignoring the phonetic length marking of the first vowel in *i.ni*.

yurru tjaa ini utingarr-ku-a, kurlumpu-yan ini
 man this sit/remain emu-dative-ligature pump-having sit/remain
 ‘This man is sitting with the pump (waiting) for the emu.’

The 19th century ethnographer Roth (1897: 97) was familiar with that district, and described how men imitated the ‘call’ of an emu using ‘a hollow log some 2½ feet to 3 feet long ...’, adding that ‘These “call-tubes” are met with throughout North-West-Central Queensland’. Anell’s (1960: 19) map shows reports of ‘emu-callers’ from seven locations from the south Gulf country in Queensland across to Charters Towers (and another two locations in northern NSW). The emu-caller has been likened to a cut-off didjeridu, and indeed there is a market nowadays for ones of recent manufacture, witness a Google search on the term.

So that explains what the thing is: the Kalkatungu were indeed describing a traditional device of theirs. But what of the glossing word ‘pump’?

The English word *pump* is hard to relate here semantically. But as well as the emu-caller there is one other tubular aerophone long used and made by Aboriginal people in northern Australia: the didjeridu. And in western Cape York Peninsula (CYP) it is called *pamp*, phonetically matching the English word spelled *pump*. The key is this entry in Alpher’s (1991) *Yir-Yoront lexicon*:

PAMP (N) *Etymology*: < English **bamboo**, probably via one or more other Aboriginal languages.

DIDGERIDOO. **Olo pamp palarrng**. He’s blowing a didgeridoo. *Note*: A recent cultural introduction to the area and not played at Kowanyama.

YO-PAMP (N) *SCI*: plant.

CASTOR BEAN, PALMA CHRISTI, RICINUS COMMUNIS. *Note*: Not a bamboo. *L.E.*⁵ ‘kerosene bush’.

A quite similar word is recorded in Kuuk Thaayorre, the neighbouring language to the north:

yuk **pamp** –a nn **bamboo** flute pipe (Foote and Hall 1992–1995: 101; nn abbreviates noun)

The slight differences in meaning and form are instructive. First, the denotation is a smaller aerophone, one traditionally used in the Torres Strait islands, made from a slender species of bamboo native to the Cape York region. The bamboo flute or pan-pipe apparently spread south prehistorically along western Cape York Peninsula, as the Wik languages have their own terms for it (Peter Sutton, p.c.). Also the ‘Bone or reed (blown like pan-pipe)’ was reported further south, from the southeast Gulf of Carpentaria, by the 19th century ethnographer W. E. Roth (Roth 1902 per Moyle 1967 map), so it could well be that slender bamboo substituted for other materials in the pan-pipe. In any case, it seems that when the English word *bamboo* arrived in

⁵Local English. How and when the *pamp* word came to also denote the introduced castor bean plant calls for integration into my account. Edwards and Black (1998) list the same word *yok pamp* ‘castor bean (*Ricinus communis*)’ in Kokoberrin, a neighbouring language to the south; no word is listed denoting an aerophone or bamboo. The castor bean plant and bamboo have in common that they have hollow jointed stems (as drawn to my attention by David Wilkins, p.c.) and they are fast growing, suckering, colony forming plants.

the region, it was applied in some places at least also to the pan-pipe. This may well have been prior to and thus independent of the arrival of the didjeridu in western Cape York Peninsula (CYP).

As to the form of the Kuuk Thaayorre word, Barry Alpher (p.c.) noted that loss of an earlier final V2 (i.e. the vowel at the end of a disyllabic word) is common to all these western CYP languages, and points out that the oblique form in Kuuk Thaayorre is *pampa*, with echo-vowel *a* instead of *u*; this implies that Kuuk Thaayorre heard *pamp* (rather than *pampu*) when the word was borrowed into that language. I join Alpher (p.c.) in proposing that when *bamboo* from frontier English or creole was first borrowed into a western CYP language, *pampu* was truncated to *pamp* by loan adaptation (Alpher and Nash 1999: 14–15), and the truncated form was borrowed further into Kuuk Thaayorre.

But what of Yir-Yoront *pamp*, which we are told does not comprise bamboo. Well, a possible explanation is that it came as the name of the didjeridu, when this instrument arrived, ultimately from Arnhem Land. I haven't seen an account of how this happened, but a parallel arrival in the southern Gulf has been explained this way:

The didgeridoo first entered Mornington Island in the 1930s as a result of visits to other Aboriginal settlements by the mission boat the *Morning Star*. This vessel was crewed by local tribesmen and paid occasional visits to Arnhem Land's Yirrkala mission. Items of material culture, including the didgeridoo, were brought back to Mornington Island from these visits. (iDIDJ Australia 2004)⁶

This same mission vessel's circuit also included Aurukun settlement, where lived speakers of many of the Wik languages north of Kuuk Thaayorre, but the Wik Mungkan dictionary (Kilham et al. 1986) does not record any *pamp* or didjeridu word.

In any case, in the Northern Territory the didjeridu has long been called *pampu* in Aboriginal English (with stress on the first syllable) and the word is fairly widespread especially in the northern half of the NT, and in the languages of people who adopted the didjeridu in historical times. Arthur's (1996) *Aboriginal English* has an entry for it, labelled '[northern Aust.]' and noting 'Also bamboo pipe'.⁷

The ethnomusicologist Moyle (1981: 322) remarked thirty years ago:

The fact that bamboo didjeridus were quite common among northerly groups in the Northern Territory during the last century is confirmed by the word 'bamboo' which is still used in

⁶The "interactive map shows the major areas in the 'Top End' of Australia where the didgeridoo is traditionally found", in Exhibition of Didgeridoos. Memmott (1980: 271–272) recorded the word *pampu* 'didjeridu' in oral history recalling new artefacts that came to Mornington Island with the *Morning Star* in the 1920s–30s.

⁷Somewhat surprisingly Arthur's (1996) earliest citation is as late as 1969; along with a 1957 reference to *bamboo puller* 'a didgeridoo player'. These are antedated by Worms (1953: 278): "the Arnhem Land tribes also have a 'bamboo', a sort of crude trumpet made from a narrow branch of a tree". Balfour's (1901) title applied the expression "bambu trumpets" in 1901. It might be thought that as *bambu* is a Malay word it could have been borrowed through Makassarese (Mkr) contact rather than through English; however Walker and Zorc (1981: 118) list only *bamutuka* 'pipe' < Mkr *pammudúkaŋ* 'bamboo opium pipe' root word Mkr *udu?*, Malay *udut* 'to suck-at, smoke'.

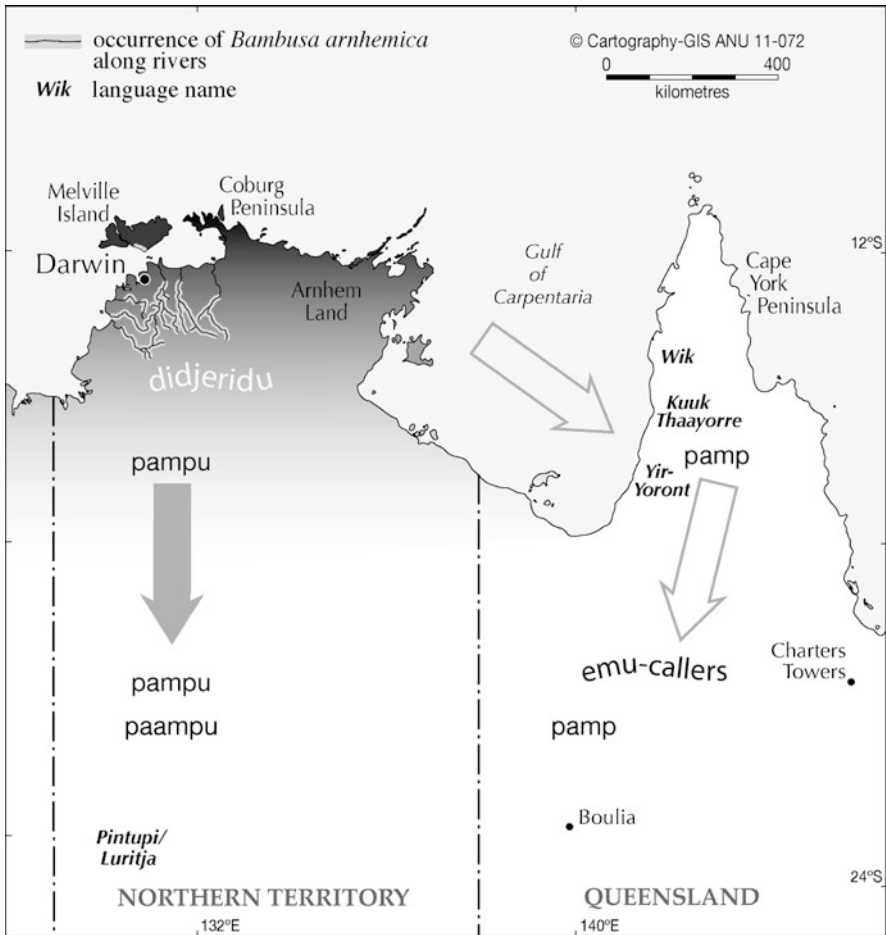


Fig. 1 Northern Australia, showing hypothesised spread of the *bamboo* word, open arrows indicate spread is inferred. Shading shows 19th century range of the didjeridu. Based on Anell (1960), Moyle (1981), and *Bambusa* range from Franklin (2008)

the lingua franca by some Aborigines when referring to the instrument, though ‘didjeridu’ may be gaining ground.

The suggestion here is that the first didjeridus were of bamboo; and that because of the availability of bamboo in the north-western region of the Northern Territory, the first didjeridu players may well have belonged to that region.

Bambusa arnhemica is the only one of the three endemic species of bamboo in Australia which is suitable for making a didjeridu. Botanists including Franklin (2008) have shown that the species is confined to Western Arnhem Land and the Daly River districts, as shown by the river lines on the map in Fig. 1.

The earliest records of the didjeridu are from this part of northern Australia, and these were made from bamboo, as noted by the earliest observers. Consider first

these quotations from the *Australian National Dictionary*'s (Ramson 1988) entry for *eboro*, a word no longer in use, nor remembered by Aboriginal people in north Australia:

1845 L. Leichhardt Jrnl. Overland Exped. Aust. 16 Dec. (1847) 534 They tried to cheer us up with their corroboree songs, which they accompanied on the Eboro, a long tube of bamboo, by means of which they variously modulated their voices.

1846 J.L. Stokes Discoveries in Aust. I. 394, I here saw the only musical instrument I ever remarked among the natives of Australia. It is a piece of bamboo thinned from the inside, through which they blow with their noses. It is from two to three feet long, is called ebroo [sic], and produces a kind of droning noise.

Prior to these writings, there are two recorded observations at the short-lived settlement 1828–29 at Fort Wellington in Raffles Bay on the Cobourg Peninsula at the far north of Australia.

The first was by naval surgeon Thomas Braidwood Wilson (Wilson 1967) who published in 1835 a book including his account of a visit to Fort Wellington. Wilson provided an illustration entitled 'Dance of the Aborigines of Raffles Bay' (Wilson 1835: 88) and described the dance 'to the music, produced by one of their part from a long hollow tube' (Wilson 1835: 87), which Kim Akerman (an expert on Australian material culture) believes is 'the earliest depiction of the didjeridu or eboro in use'.⁸ Akerman (p.c.) continued:

I think it is a bamboo one for two reasons:

1. The diameter—which is much more in proportion to the early bamboo didjeridus that I have seen, when compared with wooden ones; and
2. The lightness of the material is demonstrated by the fact that it is being held in one hand off the ground.

These points do not prove it is bamboo but I think greatly raises the possibility.

Wilson's (1835: 319) vocabulary from Raffles Bay includes

Ebero . . . Their musical instrument

which is the earliest known record of this term, predating the 1845 use by Leichhardt as quoted above from the *eboro* entry in the *Australian National Dictionary* (Ramson 1988). Note that the word has apparently dropped out of use on the Cobourg Peninsula (Bruce Birch, p.c.).

The second observation is by Captain Collet Barker (Bach 1966) who was Commandant of Fort Wellington. In his journal Barker described what we recognise as the didjeridu:

Mago had brought a kind of musical instrument, a large hollow cane about 3 feet long bent at one end. From [this] he produced two or three low & tolerably clear & loud notes, answering to the tune of didoggerry whoan, & he accompanied Alobo with this while he sang his treble. (Mulvaney and Green 1992: 113)

⁸There is a rock art image of a human figure playing a didjeridu in Kakadu National Park (Chaloupka 1993) which may well predate Wilson's drawing.

Barker’s word ‘cane’ here would not apply to a hollow limb from a tree. Note by the way his expression *didoggerry whoan*: this prefigures the word *didjeridu* for which the *Australian National Dictionary*’s (Ramson 1988) earliest citation is as late as 1919.

So, while the *didjeridu* has long been made from a variety of timbers, the association of the *didjeridu* and the bamboo plant is well established and derives from the northwestern part of the Northern Territory. The early settlements in that region (Raffles Bay, Port Essington,⁹ and then Darwin) are the plausible site of adoption of the term *pampu* among Aboriginal people, and we can infer that they learnt the word *pampu* from the early contacts with English speakers, and applied it to the *didjeridu*. The *pampu* word spread both to people who already had the *didjeridu* (and their own term for it), such as to the east in Arnhem Land, and also with the *didjeridu* itself to other people to the south who had no previous term of their own. I have indicated the southerly spread with the solid arrow on the accompanying map; as far south as for instance *paampu* “didgeridoo; from English ‘bamboo’; not used in Central Australia” in the *Pintupi/Luritja dictionary* (Hansen and Hansen 1992). The route along the Overland Telegraph Line south from Darwin has long been a conduit for the spread of innovations, and presumably was involved in particular in how Eylmann (1908: 376, Table XXIV Fig. 5) in 1897 happened to observe a bamboo ‘trumpet’ among the Warumungu of the central Northern Territory.

Another factor which might have had an influence is folk-etymology, with two aspects: (1) the English word *pump* could cover the repetitive exertion producing pulses of air, a kind of pumping; and, for the emu-caller, (2) onomatopoeia: the word *pamp* mimics the sound of the hand-struck emu-caller.

In sum, it seems that people familiar with the emu-caller adopted the ‘bamboo’ word for that somewhat similar aerophone. We can deduce that it reached the Kalkatungu via western Cape York Peninsula,¹⁰ because they adopted the truncated form *pamp* (while their language usually preserves the final vowel of a loan word). I have indicated this hypothesis by the hollow arrows on the above map.

There is another link between the two aerophones:

A suggestion that the ‘emu decoy’, reported in several parts of Australia, may have been a precursor of the *didjeridu* in some areas is to be found in an extract from Roth (1902).¹¹ (Moyle 1981: 327)

So, curiously, by spreading from the *didjeridu* to the emu-caller, the word may have reversed the course of an earlier adaptation deriving the *didjeridu* from the emu-caller.

⁹Balfour (1901: 33) includes a photograph of ‘three bambu trumpets . . . from the Alligator tribe, Port Essington’ in the Pitt Rivers Museum, Oxford. See items 1900.55.273 and 274 and 1900.71.12 in the databases at <http://www.prm.ox.ac.uk>.

¹⁰There is evidence for a parallel trade link: Sharp (1952) noted that stingray barbs from the Yir-Yoront area were exchanged for ground stone axe heads sourced a long way south, subsequently matched with quarries in Kalkatungu country (Davidson et al. 2005: 108).

¹¹Roth’s (1902: 23–24) report was from further east, from north-east Queensland.

3 *Kurlumpu*

The Kalkatungu word *kuɭumpu*, where I began, also looks to have an intriguing etymology. Being a word indigenous to Australia however its story can only be prehistoric inference. Table 1 contains the apparently cognate words I have found, cast in the source spelling or the current orthography. (Note that in the relevant Australian languages there is a single series of stop consonants, variously but not distinctively spelled with the voiced or voiceless symbols.) The words are presented in a geographic order northwest to southeast, which is also from the longer more complex form to the shorter forms.

The apparent cognates in Table 1 span a major linguistic divide in Australia, that between the Pama-Nyungan (PNy) family across most of the continent, and the disparate languages of other families across northern Australia. Given the time depth of the separation of these families and the scarcity of inherited shared vocabulary, we can presume that the cognate set in Table 1 involves borrowing, fitting with the relatively recent diffusion of the denoted artefacts, especially the didgeridu. The linguistic evidence tugs in two ways as to the direction of the diffusion: the non-Pama-Nyungan languages show more variation (initial *k* possibly lenited to *w* in Alawa; final *ng* separable in Miriwoong), whereas the *rr* segment is attested only in a Pama-Nyungan language (Warlpiri, and likely cognate with the final *t* in Mudburra). Drawing on extra-linguistic evidence, it might be thought that the word spread south (from non Pama-Nyungan languages) along with the didgeridu; but it is also possible that the word denoted a pre-existing aerophone in Pama-Nyungan languages, and was later applied to the didgeridu.

There is another apparent cognate set spanning some non-Pama-Nyungan languages of north Australia, which can be matched to the above in the first two syllables, shown in Table 2.

The words in Table 2 matching in just the first two syllables are yet to be demonstrated to be related. Similarly, for guidance of future research I mention as intriguing potential cognates some other words in Pama-Nyungan languages which are quite similar in form, but with a rather different meaning:

1. Wulguru *kulumpuru* ‘tree with honey in it’ (Donohue 2007: 41) (per Claire Bower, p.c.; Price 1885: 30: *Cooloomboro* ‘Tree “Sugar Bag” In which the natives find the Honey’); the meaning connection would be through ‘hollow tree’. The Wulguru language was spoken about 1,000 km to the east on the Queensland coast around Townsville; the *r* is Donohue’s reconstitution which matches an apical flap or retroflex glide.
2. Warumungu *kulumpurr* ~ *kulumpul* ‘hard, loud’, an eastern neighbour of Warlpiri
3. Warlmanpa *kurlumpurrnga* ‘(upper) arm’, another eastern neighbour of Warlpiri; compare also Warlmanpa’s non-Pama-Nyungan neighbour Jingulu *gurlumbu* ‘thigh bone’ (Chadwick 1975: 123). The speculative meaning connection could be through the match in size and shape between an arm and the Warlpiri trumpet.