Søren Wichmann and
Anthony P. Grant (eds.)

Quantitative Approaches to Linguistic Diversity

BENJAMINS CURRENT TOPICS 46

Quantitative Approaches to Linguistic Diversity

Commemorating the centenary of the birth of Morris Swadesh

Edited by

Søren Wichmann
MPI for Evolutionary Anthropology & Leiden University

Anthony P. Grant
Edge Hill University

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Quantitative approaches to linguistic diversity : commemorating the centenary of the birth of Morris Swadesh / edited by Søren Wichmann, Anthony P. Grant.
p. cm. (Benjamins Current Topics, ISSN 1874-0081 ; v. 46)
Includes bibliographical references and index.
1. Linguistics--Statistical methods. 2. Language and languages--Variations.
3. Mathematical linguistics. I. Wichmann, Søren, 1964- II. Grant, Anthony, 1962- III. Swadesh, Morris, 1909-1967.
P138.5.Q36 2012
410.72'7--dc23 2012027511
ISBN 978 90 272 0265 9 (Hb ; alk. paper)
ISBN 978 90 272 7335 2 (Eb)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Benjamins Current Topics

Special issues of established journals tend to circulate within the orbit of the subscribers of those journals. For the Benjamins Current Topics series a number of special issues of various journals have been selected containing salient topics of research with the aim of finding new audiences for topically interesting material, bringing such material to a wider readership in book format.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/bct

Volume 46

Quantitative Approaches to Linguistic Diversity.
Commemorating the centenary of the birth of Morris Swadesh
Edited by Søren Wichmann and Anthony P. Grant

These materials were previously published in *Diachronica* 27:2 (2010)

Preface

This volume is a reprint of *Diachronica*, volume 27, Part 2, which appeared in late 2010. The papers have been updated with regard to references and corrected for typographical errors, but otherwise appear as in the original. Most papers were presented during 17–18 January 2009 at the Swadesh Centenary Conference, which was held at the Max Planck Institute for Evolutionary Anthropology in Leipzig, with funding provided by this institution. The editors would like to thank the Director of the Linguistics Department at MPI-EVA, Prof. Bernard Comrie for supporting this event, and Claudia Schmidt for expert organizational assistance. Claudia Schmidt also helped us in the preparation of the indices for this book, along with Sabine Günther. We would also like to thank Anke de Looper and the John Benjamins Publishing Company for their alacrity in including this work in the series *Benjamins Current Topics*.

While the theme of this book, quantitative approaches to linguistic diversity, accounts for only a small part of Swadesh's work, it does signal the research arena that he has most famously contributed to, so it seems fitting to celebrate this particular theme as we celebrate his memory Thus, while reminding ourselves that Swadesh did not introduce lexicostatistics and glottochronology until 1950 when he was at the dawn of his second career and that while continuing to use these methods to the end of his career he pretty soon left it to others to develop them further, we nevertheless present papers here that mostly revolve around how to quantify aspects of language history. Although clearly in the spirit of Swadesh, most of this work could not have been carried out half a century ago for the simple reason that computers were not as ubiquitous and effective then. Swadesh's former student Daniel Cazés-Menache has told us about how computers at the Universidad Nacional Autónoma de México (UNAM) in the 1960s had to be kept cool by being housed in large rooms with blocks of ice. Some of the visions implicit in Swadesh's work are only becoming realized within the last few decades, largely because of the late advent of the personal computer.

Grant's first paper highlights some of the linguistic achievements of Swadesh outside lexicostatistics and its troubled partner glottochronology, showing how he made contributions to most if not all fields of descriptive and indeed documentary linguistics. The following papers build upon recent surges of interest in

quantitative statistical methods, mathematical and biological modelling, the study of genealogical linguistic relationships, and the ways in which these interact.

Swadesh was increasingly interested in the information which the study of the spread of language families could give about world prehistory. **Hammarström** discusses the issue of grouping and subgrouping the 7000+ languages of the world into a set of scientifically validated taxa, be they multi-member or isolates. He examines the relative sizes and directions of spread of families spoken by hunter-gatherer groups and those used agriculturalists, showing that there is a tendency for large families to pertain to groups of agriculturalists. 'Agricultural families', however, do not tend to stay within confined climatic zones as one might expect if subsistence strategy were the main factor accounting for the geospatial behaviour of language families.

The paper by **Holman** is the work of a statistician exploring the structures of linguistic phylogenies. The author examines rates at which languages become extinct or develop and split. Inspecting both trees constructed by hand and trees developed by computational methods, Holman finds that all evidence points to differences in evolutionary rates of birth and death of languages pertaining to different genealogical groups, rather than one rate applying evenly across all families.

The question of what can be borrowed (as illustrated in the World Loanword Database, a sample of 41 languages from most parts of the world) is examined in the paper by **Tadmor, Haspelmath, and Taylor**. The first two authors were the directors of the Loanword Typology Project at MPI-EVA, under whose aegis the material for the database was collected. A major result of their investigation is the Leipzig-Jakarta list, which is set up as a potential replacement of (or as a complement to) the various Swadesh lists. 100 items in length, it is composed of the forms within the database which are least frequently borrowed (or indeed never borrowed) in the sampled languages; 38 of these do not appear in the shorter Swadesh list.

The paper by **Wichmann, Müller and Velupillai** addresses a central issue of linguistic prehistory, that of how to locate the homelands of language groups. They take up the old idea in linguistics (and biology) according to which the area of highest current diversity is most probably the homeland, and operationalize this idea by measuring a 'diversity index' for each language in a language group. The language with the highest diversity index is assumed to be spoken in what was earlier the homeland, and the distribution of indices having successively lower values would then reveal directions of migration. To measure the indices, the authors draw upon geographical and linguistic distances, where the latter derive from the computerized comparison of word lists representing a reduced (40-item) version of the Swadesh list. The data were gathered under the auspices of the Automated

Similarity Judgment Program (ASJP), a consortium-driven project Involving, *inter alios*, both editors of this volume.The large coverage of this database, which is also drawn upon in the papers by Holman (see above) and Tria et al. (see below), allows the authors to hypothetically locate the homelands of as many as 82 of the world's language families, from all continents.

Grant's second paper takes a qualitative approach to lexicostatistics. He maintains that the clearest and most accurate pictures of linguistic relationships (including the identification of synapomorphies) will emerge if due attention is paid to qualitative issues, including examination of the data themselves, as a preliminary to the application of quantitative methods, and presents illustrations of this approach from several families, including Caddoan and Romance.

Heggarty's paper is part of the same approach to quantitative lexicostatistics as the Sullivan and McMahon paper introduced below. Heggarty's speciality is Andean languages, and he argues that information which may be essential for classifying languages within a family tree can be lost if one cuts down the original Swadesh list to fewer than 100 items. Lists should instead be expanded, he argues, and he demonstrates the benefits of doing this with lexical data collected in the field by Heggarty and his colleagues from Quechuan and Aymaran (Jaqi) languages. These data are used to show that Quechuan and Aymaran are unlikely to be related since lexical similarities tend to be greater for less stable than for more stable lexical concepts.

Concentrating on quantitative methods which the authors employ to investigate relationships within a subset of Germanic languages, the paper by **Sullivan and McMahon** draws upon the example set by Swadesh's work for the purposes of linguistic quantification, using phonetic criteria and applying the NeighborNet algorithm to subclassify languages.

Finally, the physicists **Tria, Caglioti, Loreto and Pagnani** introduce a new algorithm which takes distances as input for the computation of family trees. Their work represents a branch of study — general phylogenetics — which was only barely beginning to develop when lexicostatistics was first introduced. (One can only speculate about the different sort of history lexicostatistics would have had if linguists in the 1950's and 1960's had had the computational tools that are available today for creating family trees from data such as cognate counts!) The new algorithm is tested on distances calculated from ASJP word lists for a set of Indo-European languages by a method introduced by Levenshtein, and it is shown to perform well. An important novelty of the authors' method is that it allows for confidence estimates for the different branches in a distance-based phylogeny.

The papers here are united by theme and show that one aspect of Swadesh's work — that of classifying and subgrouping languages by using methods which employ various fields of mathematics — has relevance not only for descriptive and

diachronic linguists, but also for workers in fields such as biology, archaeology, and general phylogenetics. As such, this collection represents a fruitful cross-disciplinary interchange in which various fields are enhancing the unfolding picture of the prehistory of the world's languages.

Anthony P. Grant
Søren Wichmann

June 2012

Table of contents

Swadesh's life and place in linguistics

Anthony P. Grant
Edge Hill University

Morris Harry Swadesh (22 January 1909, Holyoke, Massachusetts – 20 July 1967, Mexico City) is mostly remembered now for his work on the twin fields of lexicostatistics and glottochronology, fields which he developed from 1950 to his death. But he achieved so much more in linguistics than a reputation based on his justly famous and still valuable lexicostatistical work would suggest.

The major lineaments of his life are well-known from accounts such as Newman (1967) and Hymes (1971). Morris Swadesh was raised at various places in the northeast and in the Midwest, and learned the printing trade from his father. From his parents Swadesh learned Yiddish and some Russian, and he studied Russian further on at university, in addition to French and German. (Incidentally he must be one of an elite set of people to have had a conversational knowledge of all six official languages of the United Nations; note Swadesh et al. 1966, and Swadesh 1948.) He attended the University of Chicago, majoring in languages, and went to study at Yale under the linguist Edward Sapir, to whose multifaceted intellectual legacy Swadesh was the major heir. He conducted fieldwork through the 1930s, worked on a Tarascan (P'urhépecha) language and literacy project for the Mexican government, setting up a newspaper in Tarascan, and then worked as a linguist under Henry Lee Smith at the Foreign Services Institute for the US Army during World War II, and gained a three-year Guggenheim Fellowship after the war. Denied academic positions in the US because of McCarthyite scares and his avowed leftist sympathies, he worked on a small budget from 1949 to 1956 (the era in which his major publications on lexicostatistics and glottochronology appeared) and spent the remainder of his life in Mexico. He taught at some of the most distinguished universities, enthusing generations of Mexican linguists with a view of linguistics inspired by (but going well beyond the confines of) American structuralism.

Arguably Swadesh's major legacy is intangible; it is to be found in the careers of his students and thereby their students. Despite being unable to teach in American universities throughout the 1950s (and having had his passport impounded for

much of this period, being therefore able only to visit Mexico), he trained and encouraged a number of students and early-career linguists, especially in Mexico, training people whose fame is secure in Middle America (and marrying one of them, Evangelina Arana, with whom he worked in Mexico and Ghana). Most notably in the USA he advised Floyd Lounsbury, a giant of anthropological linguists, and is responsible for making the linguistic insights of American structuralism known in Mexico. Swadesh and Lounsbury worked together only briefly; Swadesh had begun the Oneida Language Project, working with a tribe of Iroquoian-speaking Indians around Green Bay, Keshena and DePere, Wisconsin, an enterprise funded by the Works Projects Administration of the Franklin Delano Roosevelt administration, with the purpose of documenting the stories and customs of the Oneidas in their own language and using an orthographical system which could be used to write Oneida with optimal efficiency. Having started the project up Swadesh soon found that he had to pass it on to someone else. His replacement was a graduate student in mathematics and native Wisconsinian, Floyd Lounsbury, who later went on to revolutionise our understanding of how complex and so-called polysynthetic languages such as Oneida worked, as well as increasing our understanding of the hieroglyphic writing system of the ancient Maya civilisation, and also of the ways in which societies classify the biological and marital relationships which people have in 'kinship systems'. Swadesh had chosen well in picking his successor. He and Lounsbury did collaborate on one project, however. Many of the Oneidas had been converted to Protestantism in the 19th century, and sang hymns (either those translated from English or original compositions) in Oneida. At the Oneidas' request and as a quid pro quo for the folklore and narrative texts which the Oneidas provided Lounsbury, he and Swadesh — the lapsed Lutheran and the secular Jew — produced and printed a collection of Oneida-language hymns with the inestimable help of Oscar Archiquette, an Oneida-speaker who swiftly became literate in the orthography for Oneida which they had designed (Swadesh, Lounsbury and Archiquette 1941). In producing such work involving written production of Native text with Native concerns by the native speakers of a language Swadesh was following in the footsteps of his mentor Edward Sapir.

But much of Swadesh's legacy and his view of linguistics is accessible to modern scholars in terms of his extensive library of published works (and in the form of his fieldnotes on dozens of languages, preserved at the Library of the American Philosophical Society in Philadelphia). Swadesh collected material on a number of Native North American languages which were at best fragmentarily remembered when he was working on them (such as Biloxi, Atakapa, Mahican, Miluk Coos) and provided crucial documentation of over a dozen languages of California and the Pacific Northwest, most of them associated with the Penutian hypothesis (many of which are now dormant), providing almost the only known sound recordings of

some of them. Additionally, operating east of the Continental Divide, he worked on Catawba of South Carolina (another now dormant language), and did even more extensive work on the Louisiana isolate Chitimacha (his documentation of this language is still unpublished in its original form). Largely as a result of his field experiences while working on highly endangered languages, which were often in a state of severe structural attrition, he wrote an insightful and still important paper (Swadesh 1948) about language endangerment and what could (and had) to be done about it, in which he adumbrates almost all the issues facing people concerned with endangered languages today.

Not all Swadesh's linguistic work, though, was in the form of salvage linguistics. Swadesh worked on the language, literature and anthropology of several Native American groups whose languages were then in everyday use, working especially closely with the speakers of Nuuchahnulth (formerly known as Nootka), but he also worked with groups such as the Nez Perces, the Malecites, Penobscots, Menominees and Potawatomis, and also, in the 1960s, with the Gur-speaking Mamprusi people in Ghana (Swadesh & Arana 1967).

His brief period of work on Yupik Eskimo (written up in Swadesh 1951–1952) is illustrative of how incisive Swadesh's linguistic work could be. It was done in 1936 with a young Yupik speaker from the village of Unaaliq named James Andrews, who had come to Yale to take part in a sporting event, the Sportsmen's Show, and Swadesh spent just a few hours collecting data, because these were all Mr. Andrews could spare. But he got enough data to give his readers a good idea of the grammar of the language, to document some 500 roots, which is almost a quarter of the separate roots which occur in the vocabulary of the language, enough evidence to show that there are at least two different Eskimoan languages in Alaska, and not one as had previously been suspected. Swadesh was also able to furnish us with a short text which details Mr. Andrews' adventures as he made the long journey from the canning plant where he worked in Alaska over to Connecticut in order to take part in the sporting competition. Swadesh was one of the first to document basic divisions within 'Eskimoan' languages, recognising that (for instance) Inuktitut and Central Alaskan Yupik belong to different groups within Eskimoan.

Swadesh emulated Franz Boas and Edward Sapir in their belief in the importance of teaching native speakers how to write their own languages by setting up literacy programs in Mexico, most notably, as previously mentioned, for the linguistic isolate Tarascan — a project for which he learned to speak and write both P'urhépecha and Spanish.

Swadesh believed in popularizing linguistics and disseminating its concepts to a wider audience. He worked on materials to assist people in learning a range of languages quickly (producing work for Burmese, Russian, Arabic, and most fully, for Mandarin Chinese), and among his numerous articles are various entries

about language written for *Collier's Encyclopedia*, newspaper articles about language, and other popular works.

Some people found Swadesh's exuberance indigestible (his writings from this time are short of malice but they can be melancholic), but even in extremis he still had friends in all sorts of places, be they Ivy League universities, Mexican villages or Indian reservations. Much to the unjustified outrage of many other professional linguists, he got on with (and took seriously) members of a leading Protestant evangelical and linguistic organisation, the Summer Institute of Linguistics or SIL and their work, and even contributed an article to a volume commemorating the achievements of SIL's founder William Cameron Townsend (known as 'Uncle Cam' to some; Swadesh 1961), and wrote the preface to SIL member Elliott Canonge's collection of Comanche texts (Canonge 1958).

Evidence of such support for Swadesh in his unhappier times from an unusual quarter was published in *Current Anthropology* in 1995 by Jay Powell. This is a letter of support written by the Aht Band of Nootka Indians, in English and Nuuchahnulth (using a spelling system which Swadesh taught them in evening classes on Vancouver Island) calling for his reinstatement, and stating (erroneously) to the reader that he must have been sacked because he showed respect to Indians, African-Americans and other less privileged people.

Swadesh provided crucial input into several subfields of linguistics, especially in phonemic and morphophonemic theory (he was one of the first people to speak and write about 'morphophonemics', and did important work on English syllabics). With several Yale colleagues, such as his first wife Mary Haas (their early marriage did not survive but they remained good friends and Swadesh was to visit Berkeley frequently), he helped streamline American phonemic transcriptional practices, helping in the development of the transcriptional system now known as "American Phonemic". He wrote well-informed papers on most if not all the major branches of linguistics: phonetics, phonology, morphology, syntax, semantics, lexicology, sociolinguistics, psycholinguistics, historical and comparative linguistics and linguistic anthropology.

Swadesh's work, which is always clear and readable, combines Sapirian vision and Bloomfieldian precision (including a Bloomfieldian fondness for collecting and working with text corpora, as his Chitimacha work indicates). To highlight a branch of his work which is too little-known in the Anglophone world because it appeared in Spanish, his work on a number of indigenous Mexican languages (of which Swadesh & Sancho 1966 on Classical Nahuatl is best known, though Tarascan, Classical Yucatec and Classical Mixtec were also tackled) shows that he was not afraid to tackle the totality of a language, including its possession of an extensive vocabulary. His procedure here was to restate it (a popular practice among American Structuralists of Swadesh's era) in a way which preserves all the material

while presenting it in a concise format, but also one which enables the reader to understand and use this in order to gain a fuller appreciation of original sources on the language. Consequently it is deplorable that so little of his work has been reprinted. A volume entitled *Selected Works of Morris Swadesh* would provide a solid linguistic and (in the best senses) humanistic education.

Though a man obsessed by the science of language, Swadesh was adept at using findings from other fields of knowledge to assist in his linguistic researches. His work on glottochronology, inspired as it is by Willard Libby's work on carbon-14 dating and by Douglass' work on dendrochronology, exemplifies this. But he also did early work (Swadesh 1963) on setting up a notational linguistic transcriptional system for using computers for processing the material necessary for large-scale comparative linguistic work; his techniques have been followed by others, such as Terrence Kaufman (for instance Kaufman 1971), and similar techniques of notation are currently being employed by the Automated Similarity Judgment Project (Brown et al. 2008) which is discussed in Wichmann et al. and Holman et al. (both this volume). The interaction of different fields of science in Swadesh's work was bidirectional: Swadesh used historical insights, including lexicostatistics, in an attempt to cast more light on aspects of the history of Mesoamerica, and indeed on world linguistic prehistory (as Hymes 1971, a biographical coda to Swadesh's last, most ambitious, and regrettably unfinished book, makes clear).

Swadesh published over 230 books, articles, chapters, reviews, notes and abstracts, written in English and later sometimes in Spanish, and touching on Old and New World languages (Hymes' 1971 article is followed by a complete bibliography of Swadesh's works). His work draws incessantly on wide reading, on a sound knowledge of all the branches of linguistics (many of which he helped develop) and on deep personal experience of working on or learning over twenty languages. Morris Swadesh was a versatile linguist in both popular and technical senses. Had Swadesh's work attracted more than a fraction of the attention that certain other linguists' work attracted in the 1950s and 1960s, our understanding of the workings of many more languages would have been far greater. But one who reads Swadesh's work comes to realise that despite its vastness, the world of languages can be seen more clearly through its thousands of language systems. And after reading Swadesh's work on languages such as Nahuatl or Chitimacha we can appreciate the myriad facets of the world inside each language.

References

Brown, Cecil H., Eric W. Holman, Søren Wichmann, & Viveka Velupillai. 2008. "Automated classification of the world's languages: A description of the method and preliminary results". *STUF — Language Typology and Universals* 61:4.285–308.

Canonge, Elliott. 1958. *Comanche Texts. Summer Institute of Linguistics Publications in Linguistics*, 1. Norman: Summer Institute of Linguistics of the University of Oklahoma.

Hymes, Dell. 1971. "Morris Swadesh: From the Yale school to world prehistory". *The Origin and Diversification of Language* ed. by Morris Swadesh & Joel Sherzer, 228–270. New York: Aldine.

Kaufman, Terrence. 1971. "Areal linguistics and Middle America". *Current Trends in Linguistics* 11.459–483.

Newman, Stanley. 1967. "Morris Swadesh". *Language* 43.948–957.

Powell, Jay. 1995. "To see ourselves as others see us". *Current Anthropology* 36.661–663.

Swadesh, Mauricio, María Chuairy & Guido Gómez. 1966. *El árabe literario*. México, D.F.: El Colegio de México.

Swadesh, Morris. 1948. "Sociologic notes on obsolescent languages". *International Journal of American Linguistics* 14.226-235.

Swadesh, Morris. 1951–1952. "Unaaliq and Proto-Eskimo". *International Journal of American Linguistics* 17.66–70, 18.25–34, 69–76, 166–171, 241–256.

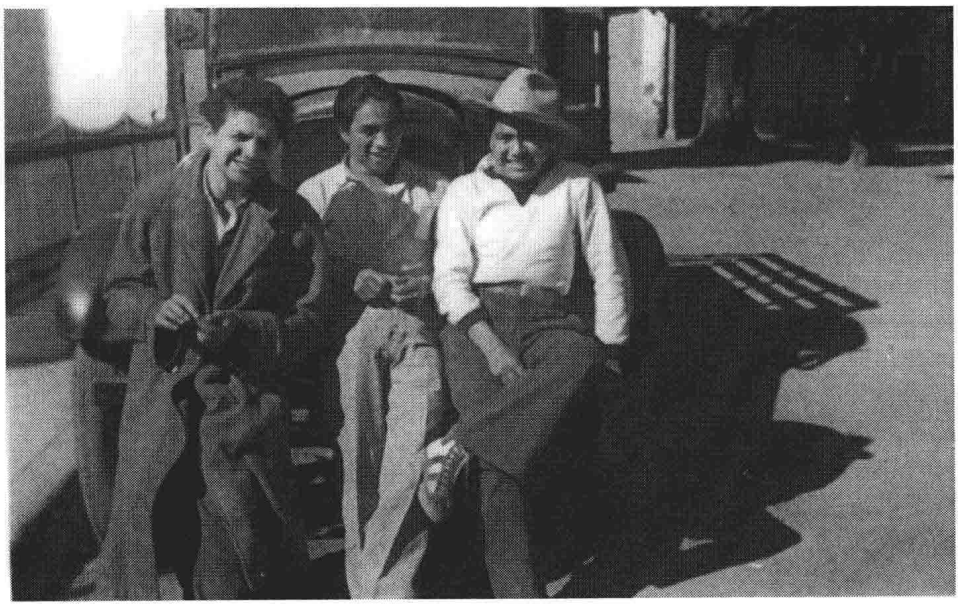
Swadesh, Morris. 1961. "The culture-historic implications of Sapir's linguistic classification". *A William Cameron Townsend en el vigesimoquinto aniversario del Instituto Lingüístico de Verano,* 663–671. México, D.F.: Instituto Lingüístico de Verano.

Swadesh, Morris. 1963. "A punchcard system of cognate hunting". *International Journal of American Linguistics* 20.283–288.

Swadesh, Morris & Evangelina Arana. 1967. *Diccionario analítico de la lengua mampruli*. México, D.F.: Museo de las Culturas.

Swadesh, Morris, Floyd Lounsbury & Oscar Archiquette. 1941. *OnAyoda'a:gá' deyelihwahgwá:ta*. Oneida, WI: Oneida Tribe of Wisconsin.

Swadesh, Morris & Madalena Sancho. 1966. *Los mil elementos del mexicano clásico*. Mexico, D.F.: Universidad Nacional Autónoma de México.



Morris Swadesh with two P'urhépecha (Tarascan) speakers, Leopoldo Hernández Cruz (from Naranja) and Ruben Salvador (from Tarejero), in Michoacán, Mexico in 1939 (reproduced courtesy of a member of the Swadesh family)

A full-scale test of the language farming dispersal hypothesis

Harald Hammarström
Radboud Universiteit Nijmegen and Max Planck Institute for Evolutionary Anthropology

One attempt at explaining why some language families are large (while others are small) is the hypothesis that the families that are now large became large because their ancestral speakers had a technological advantage, most often agriculture. Variants of this idea are referred to as the Language Farming Dispersal Hypothesis. Previously, detailed language family studies have uncovered various supporting examples and counterexamples to this idea. In the present paper I weigh the evidence from ALL attested language families. For each family, I use the number of member languages as a measure of cardinal size, member language coordinates to measure geospatial size and ethnographic evidence to assess subsistence status. This data shows that, although agricultural families tend to be larger in cardinal size, their size is hardly due to the simple presence of farming. If farming were responsible for language family expansions, we would expect a greater east-west geospatial spread of large families than is actually observed. The data, however, is compatible with weaker versions of the farming dispersal hypothesis as well with models where large families acquire farming because of their size, rather than the other way around.

1. Introduction

Some language families are 'large', like the Austronesian family, which is both geographically widespread (from Madagascar to Easter Island) and have a large number of member languages. Yet other families are minimal in both geographic size and in the number of member languages, e.g. isolates like Burushaski. A natural question is: why are some small when others are large?

One attempt at explaining why some language families are big (while others are small) is the hypothesis that the families that are now large became large because their ancestral speakers had a technological advantage, most often that of

farming. In this paper, I focus on this hypothesis, to see how well it accounts for surface properties of the language families in the world. I have developed a database of *all* language families and approximations of their cardinal size (number of member languages), geospatial size and subsistence type (i.e., whether speakers have a predominantly hunting-gathering or a farming subsistence). This database enables us to perform a number of statistical tests of hypotheses involving size and subsistence type of language families. In this way, some merits of farming-dispersal hypotheses that were previously opaque on a worldwide scale are elucidated.

2. Language families and data

2.1. Language families

There are some 7,000 attested languages in the world (see Lewis 2009 for a fair catalogue of the living ones). A language family is defined as follows:

- a **set of languages** (possibly a one-member set)
- with at least one **sufficiently attested** member language
- that has been **demonstrated in publication**
- to **stem from a common ancestor**
- by **orthodox comparative methodology** (Campbell & Poser 2008)
- for which there are **no** convincing published attempts to demonstrate **a wider affiliation.**

I know of no dedicated effort at a systematic application of this definition that spans the whole world. Therefore, I have used my own attempt in the present paper, yielding ca. 400 families for the 7,000 languages. A list with sources and brief commentary is given as an online appendix. Discussion of the most accurate assessment of language families is beyond the scope of this study.

A more fine-grained test could be done with a classification into both families and lower-level subfamilies within families. However, much more work would be needed for systematic subgrouping of the world's language families than for mere family membership. At this time, reliable information of this kind is not within reach for most of the world's families.

The concept of size has different senses, and there are different measures of the size of a language family. In this study, I use two measures, cardinal size and geospatial size.