CITIZEN SCIENCE IN LINGUISTICS:

PAST, PRESENT AND FUTURE

Sabine Stoll University of Zurich Department of Comparative Linguistics





European Research Council Established by the European Commission









Tracing the roots of syntax with Bayesian phylogenetics

Luke Maurits¹ and Thomas L. Griffiths

Department of Psychology, University of California, Berkeley, CA 94704

Edited by Richard N. Asin, University of Fochester, Fochester, WY, and approved July 25, 2014 (received for review October 8, 2013)

The ordering of subject, verb, and object is one of the fundamental components of the syntax of natural languages. The cistribution of basic word orders across the world's languages is highly nonuniform, with the majority of languages being either subjectobject-verb (SOV) or subject-verb object (SVO). Explaining this fact the explanatory adequacy of any putative theory of word-order functionality.

Some authors have suggested that all present larguages are descended from a common ancestor with SOV word order (10, 11). This hypothesis is interesting in light of results suggesting

$\mathsf{H} \mathsf{I} \mathsf{I} \mathsf{H} \mathsf{H} \mathsf{K}$

SANG

doi:10.1038/nature09923

Evolved structure of language shows lineage-specific trends in word-order universals

Michael Dunn^{1,2}, Simon J. Greenhill^{3,4}, Stephen C. Levinson^{1,2} & Russell D. Gray³

Evolutionary dynamics of language systems

Simon J. Greenhill^{a,b,1}, Chieh-Hsi Wu^c, Xia Hua^d, Michael Dunn^e, Stephen C. Levinson^{f,g}, and Russell D. Gray^{b,h}

^aARC Centre of Excellence for the Dynamics of Language, Australian National University, Canberra, ACT 0200, Australia; ^bDepartment of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, 07745 Jena, Germany; 'Department of Statistics, University of Oxford, Oxford OX1 3LB, United Kingdom; ^dMacroevolution and Macroecology, Division of Ecology, Evolution, and Genetics, Research School of Biology, Australian National University, Canberra, ACT 0200 Australia; eDepartment of Linguistics and Philology, Uppsala University, 75238 Uppsala, Sweden; Max Planck Institute for Psycholinguistics, 6525 XD, Nijmegen, The Netherlands; ⁹Comparative Linguistics, Radboud University Niimegen, 6525 HP, Niimegen, The Netherlands; and ^hSchool of Psychology, University of Auckland, Auckland, New Ze

Edited by Tomoko Ohta, National Institute of Genetics, Mishima, Japa

Language Phylogenies Reveal **Expansion Pulses and Pauses in Pacific Settlement**

R. D. Gray,¹ A. J. Drummond,² S. J. Greenhill¹

Structural Phylogenetics and the **Reconstruction of Ancient**

Language History

Michael Dunn,^{1*} Angela Terrill,^{1,2} Ger Re Robert A. Foley,³ Stephen C. Levinso

SCIENCE www.sciencemag.org



Nicole Creanza^a, Merritt Ruhlen^b, Trevor J. Pemberton^c, Noah A. Rosenberg^a, Marcus W. Feldman^{a,1}, and Sohini Ramachandran^{d,e,1}

NAS

^aDepartment of Biology and ^bDepartment of Anthropology, Stanford University of Manitoba, Winnipeg, MB, Canada R3E 0J9; and ^dDepartr **Cortical tracking of hierarchical linguistic structures** Biology, Brown University, Providence, RI 02912

variation in human populations

in connected speech This contribution is part of the special series of Inaugural Articles by

Nai Ding^{1,2}, Lucia Melloni³⁻⁵, Hang Zhang^{1,6-8}, Xing Tian^{1,9,10} & David Poeppel^{1,11}

Language is not isolated from its wider environment: Vocal tract influences on the evolution of speech and language

Dan Dediu^{a,b,*}, Rick Janssen^a, Scott R. Moisik^{a,c}

nature human behaviour DOI: 10.1038/s41562-017-01

Grammars are robustly transmitted even during the emergence of creole languages

Damián E. Blasi^{1,2*}, Susanne Maria Michaelis^{2,3} and Martin Haspelmath^{2,3}



Sound–meaning association biases evidenced across thousands of languages

Damián E. Blasi^{a,b,c,1}, Søren Wichmann^{d,e}, Harald Hammarström^b, Peter F. Stadler^{c,f,g}, and Morten H. Christiansen^{h,i}

^aDepartment of Comparative Linguistics and Psycholinguistics Laboratory, University of Zürich, CH-8006 Zurich, Switzerland; ^bDepartment of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, 07745 Jena, Germany; "Discrete Biomathematics Group, Max Planck Institute for the Science of Human History, 07745 Jena, Germany; "Discrete Biomathematics Group, Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany; ^dUniversity of Leiden, 2311 BV Leiden, The Netherlands; ^eKazan Federal University, Kazan, Russia,

RESEARCH ARTICLE

The Neurophysiology of Language Processir Shapes the Evolution of Grammar: Evidence from Case Marking

Balthasar Bickel¹*, Alena Witzlack-Makarevich², Kamal K. Choudhary³, Matthias Schlesewsky^{4,5}, Ina Bornkessel-Schlesewsky^{5,6}

OPEN CACCESS Freely available online

PLOS ONE

Is "Huh?" a Universal Word? Conversational Infrastructure and the Convergent Evolution of Linguistic Items

Mark Dingemanse¹⁺, Francisco Torreira¹, N. J. Enfield^{1,2}

11 anguage and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, 2 Centre for Language Studies, Radboud University Nijmegen, The Netherlands

Kinship Categories Across Languages Reflect General **Communicative Principles**

Charles Kemp¹* and Terry Regier²

www.sciencemag.org SCIENCE VOL 336 25 MAY 2012





HUGE UNRESOLVED QUESTION:

How can language be acquired by children ?

EXTREME LINGUISTIC VARIATION



EXTREME CULTURAL VARIATION



EXTREME INDIVIDUAL VARIATION



Source: Wordbank

WE NEED:

1. Data on as many diverse languages and cultures as possible

2. Data on as many children as possible recorded in their natural context



SOLUTION: CITIZEN SCIENCE





BIG CITIZEN SCIENCE PROJECT IN A SMALL COMMUNITY

IN THE FIELD: LANGUAGE ACQUISITION IN CHINTANG







Map data ©2018 Google

LINGUISTIC ANALYSIS

lexicon

orthography development

grammar analysis

with citizens





DATA PREPARATION



with and by citizens



ANNOTATIONS

grammatical annotations

situational annotations

gestural annotations

by linguists



| AT - | | dra diges appens diges appens tel to tel hande hande mens person tel tel tel tel tel tel tel tel | | 0041142 0041152 0044152 0044152 0044202 0044202 0044202 0044202 0044202 0044202 0044202 0044202 0044102 0044102 | 10 0041144.800 0000180.7 10 0041154.800 0000180.7 124 004121.800 0000180.801 144 004121.840 0000180.7 100 014121.840 0000180.801 100 014124.840 0000180.5 100 014124.840 0000180.5 101 014128.240 0000180.5 102 014128.240 0000180.5 103 014128.240 0000180.5 1041 014128.240 0000180.5 105 014128.240 0000180.5 1041 014130.800 000318.01 105 014120.880 003818.01 106 014120.880 003818.01 108 014130.800 003818.01 |
|------------------|----------------------------------|--|--------------------------|--|---|
| | | | Selector Note 🗌 Long Mol | | |
| | | | | | |
| - Installation 1 | 00 0041-16-000 0041-17-000 00-41 | 18.000 00.4119.000 00.4120 | 000 004121000 0 | 14122.000 004123.000 004124.000 | 004125300 004125 |
| 10.07 | CLDCADRIDIGH, 1022 | DTDOPHINE | CLLDCNIRS28 | CLUDCHORODOM, 1024 | 121100 |
| - 10.17 | e dujena optena | Aanchi Ib | kanchi | to hanako hanako mama | P1 |
| TRAIT | a dujana jogda na | Aanchi do | hanchi | to hanako hanako mama | |
| anna a | a dul par copt a na | Ranchi Ito | kanchi | to han to han to mame | - |
| - 558.47 | RHOLA MO -CL MM -MP PTOL | younge_OEM.U | promposition at | DENLIP in -GE in -GE methods | P704 |
| - 38.0 | C N N C C CN | CN C | CN | C C C C C A | ¢ |
| - 19.11 | | | | | |
| Lean | | | | | |
| 100.07 | PariLook at the two people. | There, Kanshi | Kanchi. | There, your uncle, your uncle. | Report |
| 1000 | a) gi annai (r a : | - 1 mail - | weed - | tel frei) frei) ete ete : | 54.7 |
| 100.07 | | | | | |
| Conversion of | | | | | |
| - 58-17 | 10Mar 2010 | 10Mar2010 | 10Mar(2010 | 10Mar/2010 | NUMer |
| Distant. | | | | uncoduble | |
| - giged points | | | | picture on the wall | |
| - minutes | | | | multiple prosible referents | |
| - and in state | | | | te (DEM UP), hanako manas (De-GEN net, | |
| | | | | declarative serilence | |

LINGUISTIC ANALYSIS



OUR DILEMMA



Corpus size: 250 h of recording

THE CHINTANG TEAM AND FUNDING

PIs Balthasar Bickel & Sabine Stoll

Banjade, Goma Bibiko, Hans-Jörg Bickel, Balthasar Berthold, Falko Cannedo, Laura Domberg, Andreas Erlach, Michael Ghimire, Sandesh Gaenszle, Martin Hegetschweiler, Shirin Harbodt. Silke Kantak, Nicole Klein, Felix Krämer, Tatjana Labs, Kristina Lieven, Elena Lorenz, Nicole Paudyal, Netra Pettigrew, Judith Polkau, Claudia Poppitz, Sindy Prokopchuk, Joel

Rai, Chandra Rai, Anita Rai, Daya Rai, Durga B. Rai, Durga K. Rai, Ganesh Rai, Janaki Rai, Lash K. Rai, Rikhi Rai, Novel K. Rai, Shanti M. Rai, Manoj Rai, Rabindra Sanjuan, Marcel Sauppe, Sebastian Schackow, Diana Schikowski, Robert Schneider Sarah Seeg, Jenny Strassberg, Joel Stoll, Sabine Vissiennon, Kodjo

Washington, Kerstin Weymuth, Rachel Wienholz, Anna Wohlgemuth, Jan Wolters, Kristin Wunder, Saskia Zakharko, Taras Zeiske, Mathias

DFG Deutsche Forschungsgemeinschaft EUROPEAN SCIENCE FOUNDATION



Fonds national suisse Schweizerischer Nationalfonds Fondo nazionale svizzero Swiss National Science Foundation







Situation now

Envisioned future situation



LINGUISTIC ANALYSIS



BUT HOW TO GET REALLY BIG DATA ?

MECHANICAL TURK

Looking to work on tasks? Sign in as a Worker | Learn more



Amazon Mechanical Turk (MTurk) operates a marketplace for work that requires human intelligence. The MTurk web service enables companies to programmatically access this marketplace and a diverse, on-demand workforce. Developers can leverage this service to build human intelligence directly into their applications.

CITIZEN SCIENCE PROJECTS





Shakespeare's World Project:

transcribe handwritten documents by Shakespeare's contemporaries

CITIZEN SCIENCE PROJECTS



Anzahi Korrekturen: D

Project SAC-KOKOS:

Eine Schweizerreise eines Gelehrten im XVI. Jahrhundert.

Jahrbuch 1883-mul

Von Prof. Dr. G. Meyer von Knonau (Section Uto).

14

#5 P1

437 /791 >

3 B AT.

Das XVL Jahrhundert ist auch für die Schweiz, ganz abgesehen von der geistigen Bewegung der Ke-formation, eine Zeit frischen geistigen Erwachens OPH Nachwirkungen der

| Em | • × × | | mus, theils in enger des religiösen Lebens eine erstmalige der schweizerischen nan nicht diese | | |
|--------------------|-----------------------------------|-------------|--|--|--|
| Anl | % $%$ $%$ $%$ á ç as ce à e 6 , * | | | | |
| erw wis: Ges | Ke-formation, | | | | |
| wis | 🗑 Löschen | 🖌 Speichern | schlechthin mit der | | |
| Net | | | er Kirche verbinden: | | |

Eine Schweizerreise eines Gelehrten im XVI. Jahrhundert.

Seite ist fertig korrigiert

Prof. Dr. G. Meyer von Knonau (Section Uto).

Das XVI. Jahrhundert ist auch für die Schweiz, ganz abgeschen von der geistigen Bewegung der Reformation, eine Zeit frischen geistigen Erwachens gewesen, und theils unter den Nachwirkungen der Errungenschaften des Humanismus, theils in enger Anlehnung an die Neugestaltung des religiösen Lebens erwuchs in dieser Zeit eine erstmalige wissenschaftliche Bearbeitung der schweizerischen Geschichte. Zwar darf man nicht diese wissenschaftlichen Leistungen schlechthin mit der Neubildung auf dem Boden der Firshe worbinden ; denn neusde der min mellen bien

an online platform to manually correct OCR errors



Institute of Computational Linguistics

GAMIFICATION in the LANGUAGE SCIENCES – a longstanding enterprise –

The Wolf and the Crane (Aesop)

A Wolf had been gorging on an animal he had killed, when suddenly a small bone in the meat stuck in his throat and he could not swallow it. He soon felt terrible pain in his throat, and ran up and down groaning and groaning and seeking for something to relieve the pain. He tried to induce every one he met to remove the bone. "I would give anything," said he, "if you would take it out." At last the Crane agreed to try, and told the Wolf to lie on his side and open his jaws as wide as he could. Then the Crane put its long neck down the Wolf's throat, and with its beak loosened the bone, till at last it got it out.







Done

GAMIFICATION in the LANGUAGE SCIENCES – a longstanding enterprise –

































CITIZEN SCIENCE PROJECTS AND GAMIFICATION







Project Discovery

help scientists to identify human proteins within human cells

MY VISION

A digital laboratory to research child-caretaker interactions in a quasi-naturalistic environment

Collaboration with:





ATTILA SZANTNER



Z



FUTURE PLANS: DIGITAL LAB



Thank you very much !







Established by the European Commission