An Introduction to Genetics for Language Scientists

During the last few decades we have discovered an enormous amount about our genomes, their evolution and, importantly for linguists and language scientists, the genetic foundations of language and speech.

Accessible and readable, this introduction is designed specifically for students and researchers working in language and linguistics. It carefully focuses on the most relevant concepts, methods and findings in the genetics of language and speech, and covers a wide range of topics such as heritability, the molecular mechanisms through which genes influence our language, and the evolutionary forces affecting them.

Filling a large gap in the literature, this essential guide explores relevant examples including hearing loss, stuttering, dyslexia, brain growth and development, as well as the normal range of variation. It also contains a helpful glossary of terms and a wide range of references so the reader can pursue topics of interest in more depth.

DAN DEDIU is Senior Investigator in the Language and Genetics Department at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands.

An Introduction to Genetics for Language Scientists

Current concepts, methods and findings

Dan Dediu

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107001299

First published 2015

Printed in the United Kingdom by TJ International Ltd., Padstow, Cornwall

A catalogue record for this publication is available from the British Library

ISBN 978-1-107-00129-9 Hardback

Dedicated to my mother, Maria Dediu, who managed to foster curiosity, interest in science and optimism even in the darkest of times.

Contents

Illustrations

Acknowledgements

This book is the result of a long process that began more than five years ago, with Helen Barton visiting the University of Edinburgh and convincing the young and naïve freshly minted doctor that I was to transform his PhD thesis (a monster of some 460 pages) into a readable book. The idea was that there were no introductions to genetics written for scientists interested in language and speech and the available general introductions were either too long and dense or did not really appeal to the interests of these readers.

Soon after this meeting I moved to Nijmegen, the Netherlands, joining the world-renowned Max Planck Institute for Psycholinguistics with the specific task to begin working on the genetics of language and speech, trying to build upon the enormous amount of expertise scattered across the various departments of the institute, paving the way as it were for the upcoming new department of Language and Genetics. This task was extremely interesting as it effectively meant bridging the gaps between specialties, and I worked with neuroscientists, "classical" psycholinguists, quantitative historical linguists and typologists (for lack of a better name), specialists in child language acquisition and even the odd field linguist. This taught me invaluable lessons about what is interesting, relevant and accessible to all these scientists, and made me discover topics that I did not even know existed and are potentially so important for the genetics of language and speech (such as the emerging village sign languages and the genetics of hearing deficits; thank you Connie). Likewise, teaching several introductory courses in various summer schools and giving invited lectures at conferences and workshops put me in direct contact with the maddeningly complicated and devilishly smart audience this book is designed for, and forced me to sharpen my explanations, pick the right examples and hedge my theoretical positions.

But it was no easy job. Trying to convey the importance of *differences* (as opposed to the always assumed ideal participant that one approximates by having dozens of imperfect replicas in an experimental condition) was a main stumbling block, as was dispelling the fear that what I am "really" interested in was to classify people based on their genes (no, geneticists are some of the

most vehement opponents of classifications as they experience first-hand, day-by-day the gradualness and complexity of human variation), or, equally worse, to find genetic differences that are – therefore one would say – unchangeable, reinforcing the differences between those who have and those who don't (of course, if modern genetics and evolutionary theory teaches us anything it is that the environment is *essential* to everything genetic to the point of blurring this apparently clear-cut distinction).

Of course, as any cursory look at my CV shows, I am no geneticist by training, but I have always been interested in biology (literally, I remember reading about evolution in my 4th or 5th grade thanks to my mother, a biology teacher) and I have been working on questions to do with evolution, genetics and language for the past 10 years or so. The good news is that you don't need to have had biology classes to be able to understand it, but the downside is that, as a self-taught geneticist, you will always have gaps in your knowledge. Thus, I am particularly indebted to my new colleagues – the "real geneticists", as I sometimes call them – in the Language and Genetics department here, at the Max Planck Institute for Psycholinguistics (whose work oft-times fills me with awe), especially to Prof. Simon Fisher, Dr. Sonja Vernes, Dr. Clyde Francks, Dr. Sarah Graham, Dr. Tulio Guadalupe, Dr. Nicolas Brucato, Paolo Devanna, Alessandro Gialluisi, and Amaia Carrión Castillo.

The Max Planck Institute for Psycholinguistics in Nijmegen is an amazing place for inter-disciplinary work where we somehow manage to talk across discipline boundaries and effectively collaborate with each other, a feat that is much more difficult to accomplish than one would naïvely think. It has also financially supported me through most of the time spent writing this book and its library allowed me to access all the journals, papers and books I needed to; many thanks to the librarians and especially to Karin Kastens. My current project here, the so-called "G[ɜ]bils" (Genetic biases in language and speech, funded by the Netherlands Organisation for Scientific Research [NWO] through the VIDI project 276-70-022, together with Dr. Scott Moisik and Rick Janssen) is a living example of how taking into account genetics in an inter-disciplinary framework can shed light on patterns of diversity in phonetics and phonology.

I have learned a lot from discussions with many colleagues across the years, but I want to mention in particular Prof. Stephen C. Levinson, Prof. Simon Fisher and Prof. D. Robert Ladd (University of Edinburgh, UK), who very nicely illustrate how one can successfully bridge multiple disciplines at once and yet still manage to be at the forefront of one's own scientific fields.

Several colleagues took their time to check parts of the manuscript, make many valuable suggestions and point me in the right direction, and I would like to thank Prof. Simon Fisher, Dr. Sonja Vernes, Dr. Clyde Francks, Dr. Sarah Graham, Dr. Tulio Guadalupe, Dr. Nicolas Brucato, Alessandro Gialluisi,

Amaia Carrión Castillo, Dr. Scott Moisik and Dr. Connie de Vos. I also want to thank one anonymous reviewer who had very important suggestions for improving the book, the most notable being a better coverage of quantitative genetics. Of course, any remaining errors and infelicities are mine!

This would have not been possible without the support of my wife, Dr. Alexandra Dima, who in fact got me to learn statistics in the first place and with whom I would often discuss not only data analysis strategies, but the deeper meaning and assumptions behind various statistical techniques and what *not* to read in a reported result (despite what is claimed in the abstract and conclusions), a very important skill to have. At times she encouraged (and even forced) me to continue working on this project and her suggestions and feedback made this a much better book than it would have otherwise been. Thank you so much!

A special "Dank jullie wel!" to Bernadette and Herman Arts without whom our adaptation to the Netherlands would have been so much more difficult (and we would have indubitably discovered the local cherries from Bemmel much later).

Finally, this book would have been much more difficult to write without all the *Open-Source* software I used throughout the years and I wish to thank all those who have contributed, one way or the other, to specific software packages and to the philosophy that makes this possible. My operating system has always been a flavour of `GNU/Linux`, including `Debian` (`http://www.debian.org`), `Ubuntu` (`http://www.ubuntu.com`), `ArchLinux` (`http://www.ubuntu.com`), `OpenSuse` (`http://www.opensuse.org`) and `CentOS` (`http://www.centos.org`). The book was written in LaTeX using initially `Kile` (`http://kile.sourceforge.net`) but mostly `TexMaker` (`http://www.xm1math.net/texmaker`) and `TexStudio` (`http://texstudio.sourceforge.net`) as editors/environments. All data analyses for this book used `R` (R Development Core Team, 2010), an open-source programming environment focused on statistics available free of charge at `http://www.r-project.org`. The graphs were generated using a wide variety of tools, including `R`, `LaTeXDraw` (`http://latexdraw.sourceforge.net`), `Inkscape` (`http://inkscape.org`), the LaTeX packages `pstricks`, `pst-pdgr` (pedigrees), `mhchem` (chemical formulas), `xy` (blocks and arrows diagrams), and `amssymb`, `MnSymbol`, `wasysym` and `tipa` (for various symbols including IPA), and even with `LibreOffice`'s (`http://www.libreoffice.org`) own `Draw` component.