

Corpus-based Analysis and Diachronic Linguistics

edited by Yuji Kawaguchi,
Makoto Minegishi and Wolfgang Viereck

John Benjamins Publishing Company

Corpus-based Analysis and Diachronic Linguistics

Edited by

Yuji Kawaguchi

Makoto Minegishi

Tokyo University of Foreign Studies

Wolfgang Viereck

University of Bamberg



John Benjamins Publishing Company

Amsterdam / Philadelphia



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

Library of Congress Cataloging-in-Publication Data

Corpus-based analysis and diachronic linguistics / edited by Yuji Kawaguchi, Makoto Minegishi, Wolfgang Viereck.

p. cm. (Tokyo University of Foreign Studies (TUFS), studies in linguistics, ISSN 1877-6248 ; v. 3)

Includes bibliographical references and index.

1. Corpora (Linguistics) 2. Language and languages--Variation. 3. Historical linguistics. I. Kawaguchi, Yuji, 1958- II. Minegishi, Makoto. III. Viereck, Wolfgang.

P128.C68C654 2011

410.1'88--dc23

2011045661

ISBN 978 90 272 0770 8 (Hb ; alk. paper)

© 2011 – Tokyo University of Foreign Studies

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

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

Contents

Message from the President <i>Ikuo KAMEYAMA (President, Tokyo University of Foreign Studies)</i>	1
Center for Corpus-based Linguistics and Language Education <i>Makoto MINEGISHI (GCOE Project Leader)</i>	3
Introduction <i>Yuji KAWAGUCHI, Wolfgang VIERECK and Makoto MINEGISHI</i>	7
The <i>Atlas Linguarum Europae</i> : A Diachronic Analysis of Its Data <i>Wolfgang VIERECK</i>	21
Variationism and Underuse Statistics in the Analysis of the Development of Relative Clauses in German <i>Anke LÜDELING, Hagen HIRSCHMANN and Amir ZELDES</i>	37
Variation and Change in the Montferrand Account-books (1259-1367) <i>Anthony LODGE</i>	59
Cognitive Aspects of Language Evolution and Language Change: The Example of French Historical Texts <i>Wolfgang RAIBLE</i>	75
The Importance of Diasystematic Parameters in Studying the History of French <i>Lene SCHØSLER</i>	91
The Reorganisation of Mood in the Epistemic Subsystem —The Case of French Belief Predicates in Diachronic Dynamics <i>Martin BECKER</i>	111
French Liaison in the 18 th Century—Analysis of Gile Vaudelin's Texts— <i>Yuji KAWAGUCHI</i>	133
Issues in the Typographic Representation of Medieval Primary Sources <i>António EMILIANO</i>	153
An Analysis of the Misuse of the Participle in Old Russian Texts <i>Yoshinori ONDA</i>	175
A Preliminary Analysis of Arabic Derived Verbs in the Leeds Quran Corpus —With Special Reference to Stem III (CaaCaC) <i>Robert R. RATCLIFFE</i>	189

On the Narrow and Open “e” Contrast in Santali <i>Makoto MINEGISHI, Jun TAKASHIMA and Ganesh MURMU</i>	203
The Classification of Apabhramśa —A Corpus-based Approach of the Study of Middle Indo-Aryan— <i>Tomoyuki YAMAHATA</i>	223
Changes in the Meaning and Construction of Polysemous Words: The Case of <i>mieru</i> and <i>mirareru</i> <i>Ayako SHIBA</i>	243
Language Change from the Viewpoint of Distribution Patterns of Standard Japanese Forms <i>Kanetaka YARIMIZU</i>	265
Index of Proper Nouns	285
Index of Subjects	289
Contributors	293

Message from the President

Ikuo KAMEYAMA (*President, Tokyo University of Foreign Studies*)

It was a great honor for me to participate in the international symposium entitled “Corpus Analysis and Diachronic Linguistics”. This symposium was also attended by the eight international scholars who have traveled to join us from America, England, Germany and Sweden.

I will now dwell briefly on the Global COE Program, which began in April 2007. This program is an effort by the Japanese government to strengthen its support to research and educational institutions in which internationally renowned work is taking place. The program was developed to take advantage of world-class resources to help foster the development of creative researchers who can lead in their fields, and to strengthen research and education in Japan’s centers of graduate education. In 2007, proposals were solicited in five areas. The program submitted in the area of humanities by our university was one of 12 selected nationally. The humanities category encompasses fields as diverse as philosophy, art, psychology, education and archaeology. The submission from our university was the only one selected in the area of linguistics. We believe this reflects the high level of research and education at our institution.

Our submission, entitled “Corpus-based Linguistics and Language Education” emphasizes on a field of empirical linguistics based on the uses of corpora. The program’s goal is to foster the growth of advanced researchers with international perspectives. This program continues the research conducted under the “Usage-based Linguistic Informatics” 21st Century COE Program, which concluded March 2007. The new program will build on the international joint research framework that was created over the past five years to achieve two goals, with the support of the entire university:

1. To further develop a comprehensive education program for graduate students
2. To give graduate students opportunities to perform fieldwork, build and analyze corpora, and receive language education and training, both in Japan and overseas.

I am not an expert in linguistics nor do I have a deep scholarly understanding of corpus linguistics. However, as a scholar of literature, I have a keen interest in the possibilities inherent in the field. The corpus concept was introduced into my area of specialization, Russian literature, in the late 1980s. As far as I know, this resulted in the creation of corpora for the works of authors such as Fyodor Dostoevsky and Andrei Platonov. However, it is not yet clear how effective the corpus concept will be in the development of the study of

literature. In contrast, corpus-based linguistics seeks not to use linguistic data to understand the latent properties of a text as a close system, but to understand the linguistic structure and function of a language within a larger context. So, I believe that corpus linguistics provides us with higher objectivity and richer possibilities in the field of humanities.

Still, it is my opinion that the greatest hurdles for corpus-based linguistics are still to come. Humans are creatures that cannot help but seek out meaning and possibilities of systemizing matters. It is evident that corpus linguistics has not been a field that describes only the actual uses of languages, but one that finds ways to generalize creative discoveries and to extend its insights. Its value lies in its ability to push itself. For corpus-based linguistics to grow creatively as a human science, we must help young researchers to develop innovative and unique capabilities for analysis. I believe that this is where the real importance of the current G-COE Program lies.

In conclusion, I would like, as president of this institution, to express my sincere respect to all the leading researchers who attended this symposium, for their untiring efforts. More importantly, I hope that the young scholars who attended the symposium have imbibed some of the passion that was on display, and I hope that it will help them to grow internationally competitive researchers.

July 12, 2011

Center for Corpus-based Linguistics and Language Education

Makoto MINEGISHI (*GCOE Project Leader*)

The Center for Corpus-based Linguistics and Language Education (CbLLE) was established with the express target to build an education and research center with unique strengths in the study of linguistic diversity and also in usage-based research of linguistic structure and language education. This centre builds on the strengths of the nationally high-ranking Graduate School of Area and Culture Studies of the Tokyo University of Foreign Studies (TUFS) and of the Research Institute of Languages and Cultures of Asia and Africa (ILCAA). Its educational and research uniqueness is achieved by integrating the three core areas of activities: (a) collection and analysis of naturally occurring language use data through field research, (b) compilation and analysis of large-scale corpora of language use data from a wide range of languages, and (c) application of corpus-based linguistic analyses to language education and pedagogy. A few details of the work that is being done in the above core areas follows:

Field linguistics: The field linguistics program supports fieldwork-based research on a typologically diverse set of languages, including not only the world's major languages but also lesser-studied languages. They include languages of Africa, Eurasia, and North America. It also aims at advancing typological research on the basis of the primary data from a broad range of languages. It provides a solid training to students in the methodology of collecting, processing, and analysis of the field data. The project has undertaken fieldwork-based study of a diverse range of languages of the world (lesser-studied languages in particular) and typologically-informed description of these languages. Some of the projects under this category are: Compilation of a Word List for Field Research on Khwe Languages; Field-work based study of under-studied speeches of India; Collation of Spontaneous Conversational Data of Individual Languages such as: Swahili, Russian, French, Spanish, Turkish.

Corpus linguistics: The program in corpus linguistics supports analysis of a large amount of language use data and compilation of corpora, which feed into linguistic informatics research and also into descriptive and typological research. Some of the specific targets are: Building electronic corpora and developing analysis and processing tools in order to support new ways of analyzing language data and multipurposing of the data; Developing

multilingual and multifunctional integrative corpora of language use for major languages on the basis of language use data collected in language teaching classrooms, blogs, etc.; Conducting international collaborative research and providing support in development and utilization of tools for corpus creation, morphological analysis, electronic dictionary creation, text analysis. The projects undertaken here include: Development of Electronic Dictionaries for Russian as well as Thai (separately); corpus Compilation of Data from Medium/Minor Language Groups; Development of Utility Manuals for German Corpus; Preparation of Introductory Text-book on Lexicology based on Corpus Data; Research on Corpora for Minor Language Group in EU Countries.

Linguistic informatics: The linguistic informatics program builds on the research in field linguistics and corpus linguistics components to significantly advance research in language pedagogy. It seeks to make a significant contribution to the research in language pedagogy by taking into account the factors of linguistic and cultural diversity through analysis of corpora of language use in actual contexts of language instruction, including naturally occurring conversations and learners' language use. A few studies undertaken in this context are: Research on Lexicon/language-use based on Corpora for Various Fields; Language Processing/education Technology; POS Search Engine for Spoken French as well as Spoken Spanish; Basic Research on E-learning through Moodle; Corpora of Learners' Language Use (both as an internal project as well as an international collaborative project); Creation of Language Tests based on Error Analysis of Language Use of English Learners.

The GCOE trains researchers and educators who have a clear understanding of the nature and significance of linguistic and cultural diversity and can take a flexible research approach to language structure and language education. This project equips young researchers with a broad foundation for linguistic research by providing practical trainings in field research, corpus-based research and language education. These training programs support the integrative research on linguistic and cultural diversity and usage-based linguistics by connecting effectively field data collection, data analysis, and educational application of theoretical insights obtained from the analysis.

The specific projects and tasks listed above form part of the larger plan of building an international research and education center with more generalized targets described below. The Center seeks to build a world-leading research and education center in the study of linguistic diversity and in the usage-based research of linguistic structure. The national and international infrastructure for supporting the GCoE are being built through the following activities:

Formation of an international network of collaborative research: Collaboration in corpus creation and in development of analysis tools (such as electronic dictionary systems); building a network of international collaboration and academic exchange in linguistic research and teaching within the framework of the 'Consortium for Asian and African Studies' which has its headquarters at the University.

Expansion of opportunities for academic interaction across institutions and across countries: Expanding opportunities for young researchers, as well as established scholars, within Japan and abroad to assemble and interact through visiting scholar programs and through employment.

Support program for young researchers: Providing young researchers with financial and technical support for linguistic field research, corpus creation, and education research in the field; and providing young researchers with financial support for professional development (including presentation at international conferences).

Active international dissemination research results: Building an information technological infrastructure that supports active electronic dissemination of research results; and Publishing the research results in a series of publications through international publishers that are specialized in publication in linguistics — the present volume being a contribution towards this aim.

Introduction

Yuji KAWAGUCHI, Wolfgang VIERECK and Makoto MINEGISHI

1. From dichotomy to hybrid dynamic synchrony

It is well-known that *Cours de linguistique générale*, the posthumous publication of Ferdinand de Saussure's introductory course at the University of Geneva, which advocated the dichotomy between synchrony and diachrony, emphasized the synchronic description, rather than the historical vicissitudes, of language. Riedlinger's note (dated December 17, 1909), makes it clear that synchrony took priority in Saussure's lectures.

pour se rendre compte de ce qui existe dans un état de langue, le mieux est de faire abstraction du passé. Paradoxe, mais paradoxe vrai: les signes de la langue ont leur valeur fixée par ce qui coexiste, non par ce qui précède (exemples). Godel (1969) 70

“to explain what is going on in the state of a language, the best approach is to forget the past. This is a paradox, a real paradox: linguistic signs have their value fixed by what coexists, not by what precedes them (examples).”

The prioritization of synchrony over diachrony in *Cours* probably originated from two motivations¹. First, the synchronic reality of language is directly observable and less dependent on extralinguistic effects, which better suits the conception of the immanent principle of *Cours*; in other words, “As its unique and veritable object, linguistics has the langue envisaged in itself and for itself².” Back then, this principle was essential to establish the foundation of linguistics as an independent domain of science that was based on the analysis of “langue”. Second, synchronic analysis was linked to the study of the systematic mechanism of language, which was a departure from the more psychologically oriented approach of earlier philological studies. However, some linguists insisted on differentiating between linguistic rules and laws in natural sciences, claiming that linguistic rules are no more than the results of historical constataion, because the condition of linguistic rules is unique, and they can never be reproduced in the same manner. This viewpoint, which contradicts our stance, underestimates the embryonic or ongoing changes

¹ It is beyond our scope to verify whether this endorsement of synchronic viewpoint dates back directly to *Cours*. For example, in his *Principes de grammaire générale*, Louis Hjelmslev regards H. G. Wiwel as the pioneer of synchronic linguistics.

² “la linguistique a pour unique et véritable objet la langue envisagée en elle-même et pour elle-même.”, *Cours*, 317.

that can often be observed in the same synchronic state of a language. Such an atomistic interpretation of linguistic phenomena consequently neglects the analysis of synchronic variation, since variation is identified through the recurrent occurrence of the variants in question. This holds true for the renowned French philologist Gaston Paris. He was keenly interested in the geographical variation of French, and created the post of Professor of Dialectology for Jules Gilliéron at the École Pratique des Hautes Études. However, Paris himself did not seem to indicate much interest in the ongoing variation of French.

Today, linguists would not question the existence of synchronic variation, which could be motivated regionally (diatopic variation), sociolinguistically (diastratic variation), or stylistically (diaphasic variation). In addition to the dichotomy of synchrony and diachrony, we can recognize the hybrid nature of synchrony that is referred to as “dynamic synchrony.” Such a conception of synchrony assumes that similar patterns of usage can coexist in a community during a certain period and that their mutual relations are not static but conflicting enough to give rise to a future systematic change through symptomatic synchronic variation. It is noteworthy that the appearance of large corpora of written texts for some languages has made it possible to analyze quantitatively as well as qualitatively the conditions of historical changes, not only over a long span of time, but also over a short span. This has already led to a number of studies on synchronic and diachronic variation, including the majority of the papers in this volume.³

2. *Realia* or *Fictio* in written documents

The main resources for diachronic studies are written texts, primarily. However, from the very beginning of the research, the questions of how and why these texts have existed after they were written need to be confronted. The birth of a new genre such as prose literature in Old French, for instance, was closely linked to the cognitive evolution of medieval writers. Such a cognitive change is evident in the preface of *Chronique de Pseudo-Turpin*, of which the earliest existing manuscript dates from the twelfth century.

Nus contes rimés n'est vrais. Tot est mençongie ço qu'il en dient car il n'en sievent rienz fors quant par oïr dire. (Chronique de Pseudo-Turpin, ms.BN fr.124, fol.1)

“No rhymed tale is true. All that they tell are fictions because they know them only through hearsay.”

The writer presumes that historical descriptions should not be based on hearsay, i.e. *fictio*, but should represent *realia*, the reality. Verse is not a convenient tool

³ See also Kawaguchi, Minegishi, and Durand (eds.) 2009.

for this purpose; prose is more often than not the preferable means to depict historical truth. Based on the evidence of existing documents, it would appear that medieval French writers rarely used prose before 1200. The emergence of prose was a historical event in the thirteenth century, which signaled a cognitive evolution.⁴ The oral tradition represented by verse, which the medieval intelligentsia and nobility had been accustomed to, was losing ground, and reading literature, instead of listening to it, had become the general practice in the thirteenth century.⁵ This ideological trend contributed to the creation of a new literary genre that focused on the construction of a virtual reality through written words. In the present volume, Wolfgang Raible treats this cognitive framework of medieval French texts by comparing the description of the Fourth Crusade (1202-04) by two contemporary writers, Geoffroy de Villehardouin and Robert de Clari.

Truth and reality as observed in written documents are polysemous, for they could be purely ideological, i.e. representing the writer's wish to be true to reality, or they could represent an objective truth. In Japan, it is possible, even today, to hear a number of sociolinguistically different ways of saying the same thing (as was common in ancient times) in Kabuki, the Japanese traditional theater. For instance, samurais would say, "*Itsu Edo-e mairareta?*" Monks and doctors would use "*Itsu Edo-e gozarimashita?*" Housewives would say, "*Itsu Edo-e okoshidegozarimashita?*" Prostitutes would say, "*Itsu Edo-e kiyashanshita?*" Geishas would use "*Itsu Edo-e kinasanshita?*" Further, workers would say, "*Itsu Edo-e oidenaseemashita?*" all of which mean the same thing, "When did you come up to Edo (Tokyo)?"⁶ It is really surprising that strikingly similar sociolinguistic variations can be found in the popular Japanese novels of Shikitei Samba, such as *Ukiyoburo* (1809-13) and *Ukiyodoko* (1813-14). Today, the existence of such variations at the beginning of the eighteenth century is accepted without question in Japanese linguistics. Written documents remain the most valuable and reliable sources for linguistic historians, for the analysis of dynamic synchrony of the past. Of course, there were, and still continue to be, constant disputes about the authenticity of these stylistic variations in written documents. Rebecca Posner uses an interesting metaphor to describe the insolubility of this problem.

How far literary Romance has ever been identical with the spoken dialects is a moot point: To pose the question of the relationship between the two again calls out the hare we refused to pursue earlier. Posner (1972) 49.

⁴ This historical development was not restricted to French. Also in Cornish, English and German, literature in verse preceded that in prose.

⁵ Kawaguchi 2007b.

⁶ Tanaka 1983. 173.

Linguists investigating written documents of the past are deeply indebted to the critical editing of philologists. António Emiliano, in his article in the present volume, remarks that the scholarly editing of primary sources cannot dispense with good practices and sound philological groundwork, and discusses several possible strategies concerning the typographic representation of medieval texts. Additionally, it would be better to use tagged corpora and some kind of concordancers for the linguistic analysis of historical corpora. In the present volume, Anke Lüdeling, Hagen Hirschmann, and Amir Zeldes discuss their use of four comparable deeply annotated corpora representing different historical stages of German. Martin Becker's analysis is based on the *New Amsterdam Corpus* (NAC)⁷ and the Middle French subcorpus of *Frantext*⁸. An excellent parallel in English is Matti Rissanen et al., *The Helsinki Corpus of English Texts* (1991). It is a structured multi-genre diachronic corpus and covers the period between c.730-1710. As this is an early corpus, it comes as no surprise that there are offsprings with considerable additions (and deletions) to the content, such as Ann Taylor et al., *York – Toronto – Helsinki parsed corpus of Old English prose* (2003) and Anthony Kroch – Ann Taylor, *Penn – Helsinki parsed corpus of Middle English* (2000).

Anthony Lodge and Yuji Kawaguchi use the electronic tools Loceme⁹ and AntConc¹⁰ respectively. The possibility of using such tools in linguistic analysis largely depends on the typological and grammatical characteristics of the languages under study. The development of tools useful for linguistic analysis is relatively easier for those languages that are more or less isolated and written with space, than for those that are agglutinative and written without space, like Japanese.

3. Ongoing changes in dynamic synchrony

The majority of the papers in this volume analyze ongoing changes, covering the phonetic, phonological, morphological, morphosyntactic, semantic, and pragmatic aspects of language, which appear as variation in the written documents.

Section 7 of Lodge's paper focuses on the phonetic features which characterize the North Auvergnat dialect. Spelling variation in documents represents not only dialectal but also phonological variation¹¹. In their analysis of digitized data from Bodding, Makoto Minegishi, Jun Takashima, and

⁷ <http://www.uni-stuttgart.de/lingrom/stein/corpus/>

⁸ <http://atilf.atilf.fr/frantext.htm>

⁹ <http://eserve.org.uk/loceme/>

¹⁰ <http://www.antlab.sci.waseda.ac.jp/software.html>

¹¹ Kawaguchi 2007a.

Ganesh Murmu examine whether or not the distinction between narrow “-e-” and open “-e-” is phonological in Santali, the most widely spoken language of the Munda language family. Such ongoing changes are sometimes recognized and recorded by grammarians and foreigners. For instance, S. R. Brown, a Presbyterian missionary, vividly describes “vulgar” Tokyo pronunciations in his *Colloquial Japanese* (1863), published in Shanghai.

The vulgar in Tokyo say *ai* for *ae*, and *oi* for *oe*; thus *mai*, instead of *mae*, “before”; *koi* (which means “love”), instead of *koe*, “voice.” They also often contract *ai* into a long *e*.; as *narane*: for *naranai*, “it won’t do.” But this is as bad as the dropping of the letter *h* by cockneys. Matsumura (1957) 222.

When compared to segmental variation, the variations in suprasegmental and linking phenomena are relatively less documented in written texts. Making use of the rare quasi-phonetic descriptions of Gile Vaudelin’s books, Kawaguchi analyzes the evolutionary stages of the liaison in French at the beginning of the eighteenth century.

Morphological or morphosyntactic variation is the most extensively treated kind of variation in this volume. Lodge analyzes variables in verb morphology, based on the corpus of the Montferrand Account-Books (1259-1367). The misuse of participles in Old Church Slavonic and Old Russian texts is analyzed in Yoshinori Onda’s contribution, while Robert Ratcliffe studies the phenomenon of semi-productivity in the so-called derived verb system of Classical Arabic, by using the data from the Leeds Quran Corpus. Morphological variation is sometimes closely related to ongoing syntactic change. Focusing on four main constructions of the two polysemous verbs, *mieru* and *mirareru* “to see” in the Modern and Contemporary Japanese Corpus, Ayako Shiba finds that the existential construction of *mieru* is being progressively replaced by *mirareru*. These papers deal with the variation or ongoing change in dynamic synchrony. The contributions of Lene Schøsler, and Lüdeling, Hirschmann, and Zeldes go beyond synchronic description; the models of diachronic development of the composed past tense from Latin to Modern French are demonstrated in the former, while those of the relative clauses from Old High German to Modern German are discussed in the latter.

Using the *New Amsterdam Corpus* (NAC), Becker conducts a semantico-pragmatic analysis of mood selection in Old French belief verbs, such as *cuidier* and *croire* “to believe.” He describes a diachronic conflict where the latter will expel the former, progressively infringing on the contexts dominated by the former for centuries. Finally, a corpus-based study of written documents can give us important clues about the classification of language groups. Through the corpus analysis of eight different texts of Apabhramśa, the language used in the literary works of northern India in the Middle Ages,

Tomoyuki Yamahata finds little relationship between Apabhramśa and the New Indo-Aryan languages.

4. Linguistic atlas and diachronic linguistics

Can a linguistic atlas be considered a linguistic corpus? Jean-Philippe Dalbera, the French dialectologist, is in favor of this¹². He distinguishes three different generations of geolinguistic works. The first generation establishes the concept of a linguistic atlas and its methodology. The second generation constitutes a reliable corpus of linguistic data by improving the atlas tool. He puts forward three key parameters for defining the corpus of a linguistic atlas: comparativism, diatopy, and lexicon. The analysis and use of these corpora are to be devised in the third generation. In this volume, two papers concern geolinguistic researches of the third generation as per Dalbera's definition.

The construction of the *Atlas Linguarum Europae* (ALE)¹³ that began in 1970 was without doubt the first original research attempt originating from the existing linguistic atlases of Europe. According to Wolfgang Viereck, three aspects are important with regard to the interpretation of the word maps of ALE: loanword research, etymological research, and the study of the motivations in designating certain objects. Using ALE, he deduces the religious and cultural history of Europe, which provides a magnificent macro perspective for describing the linguistic landscape or scene (*Sprachlandschaft*) of Europe on both synchronic and diachronic axes. Kanetaka Yarimizu's article is directed along a more micro perspective. His data comes from two recent dialect surveys: the *Grammar Atlas of Japanese Dialects* (GAJ), and the Glottogram survey in northern regions. Making full use of various statistical methods, he investigates the standardization processes of Japanese across five different historical stages.

5. Corpus-based analysis and diachronic linguistics

In this section, abstracts of 14 papers dealing with corpus-based analyses of different languages are presented in the order of their appearance in this volume.

1. Wolfgang Viereck, in "The Atlas Linguarum Europae: A Diachronic Analysis of Its Data," discusses the vast linguistic diversity exhibited across the European continent, with 6 language families and 22 language groups, each with a large number of individual languages. Several short-lived projects

¹² See his "Linguistic Atlases: Objectives, Methods, Results, Prospects," in Kawaguchi et al. 2007. 39-54.

¹³ Viereck (ed.) 2007.