



2013-03-14

Areal Patterns of Possessive Morphology in the Languages of Eurasia

Garrett K. Nay

Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Linguistics Commons](#)

BYU ScholarsArchive Citation

Nay, Garrett K., "Areal Patterns of Possessive Morphology in the Languages of Eurasia" (2013). *All Theses and Dissertations*. 3780.
<https://scholarsarchive.byu.edu/etd/3780>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu.

Areal Patterns of Possessive Morphology
in the Languages of Eurasia

Garrett K. Nay

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Arts

Janis Bozic Nuckolls, Chair
Deryle W. Lonsdale
David S. Eddington

Department of Linguistics and English Language
Brigham Young University
March 2013

Copyright © 2013 Garrett K. Nay
All Rights Reserved

ABSTRACT

Areal Patterns of Possessive Morphology in the Languages of Eurasia

Garrett K. Nay
Department of Linguistics and English Language
Master of Arts

The goal of this study is to confirm Eurasia as an independent linguistic area with respect to four features of possessive morphology: locus of marking, position of pronominal possessive affixes, obligatory possessive inflection, and possessive classification. Raw data on these features was taken from the WALS database and then run through an algorithm of genealogical stratification called *g*-sampling, in order to minimize the bias of the sample. The resulting *g*-units were then categorized by type and geographical area (New World vs. Old World, Eurasia vs. the rest of the world). These counts were tested for significance using Fisher's exact test.

Two features, locus of marking and possessive classification, were confirmed to be significantly different in Eurasia; the other two features were not significantly different. Possible reasons for these areal patterns—primarily structural reasons—are briefly discussed.

Keywords: linguistic typology, morphology, possession

ACKNOWLEDGMENTS

I would like to thank Janis Nuckolls for inspiring my passion for linguistic typology and providing invaluable guidance all along the way. I am also indebted to Deryle Lonsdale and David Eddington for their useful and encouraging feedback. Above all, I want to thank my wife, Eve, for her unending support and for believing in me even when I had a hard time believing in myself.

Table of Contents

List of Tables	vi
List of Figures	viii
1 Introduction: Areal Typology	1
1.1 Goals and insights of linguistic typology	1
1.2 Areal influences on linguistic features	3
1.3 The Eurasian macro-area	4
1.4 Possession as an areal feature	5
1.5 Overview of the thesis	7
2 Features of possessive morphology	8
2.1 Defining possession	8
2.2 Locus of marking	11
2.3 Position of pronominal possessive affixes	15
2.4 Obligatory possessive inflection	19
2.5 Possessive classification	20
2.6 Conclusion	23
3 Method	24
3.1 Problems with language sampling	24

3.2	A controlled sampling method	27
3.2.1	Dryer's genera-counting method	27
3.2.2	Bickel's g-sample method	28
3.3	Statistical tests	30
3.4	The present study	31
4	Results	33
4.1	Locus of marking	33
4.2	Position of pronominal possessive affixes	38
4.3	Obligatory possessive inflection	46
4.4	Possessive classification	50
4.5	Summary	57
5	Conclusions	60
5.1	The Eurasian macro-area revisited	60
5.2	Possession as a general feature	64
5.3	Limitations and possibilities	65
	References	67

List of Tables

2.1	Comparison of the two types of verb conjugation and possessive morphology in Hungarian.	18
2.2	Examples of the possessive classes of Chichimeca-Jonaz (Lastra de Suárez, 1984, pp. 24–25).	22
3.1	Example of Dryer’s genera-counting method, comparing verb-object order and noun-relative order.	28
4.1	Distribution of values for feature 24A: locus of marking in the possessive noun phrase.	34
4.2	Head-marking vs. other types (a) and dependent-marking vs. other types (b) in the New World vs. the Old World.	37
4.3	Head-marking vs. other types (a) and dependent-marking vs. other types (b) in Eurasia vs. other areas.	38
4.4	Distribution of values for feature 57A: position of pronominal possessive affixes.	38
4.5	Possessive prefixes vs. other types (a) and possessive suffixes vs. other types (b) in the New World vs. the Old World.	41
4.6	Possessive prefixes vs. other types (a) and possessive suffixes vs. other types (b) in Eurasia vs. other areas.	42
4.7	Possessive prefixes vs. other types (a) and possessive suffixes vs. other types (b) in Eurasia vs. other areas, with the genera of Sino-Tibetan reduced.	43
4.8	Distribution of values for feature 57A-r: pronominal possessive affixes (reduced).	44
4.9	Presence vs. absence of pronominal possessive affixes in Eurasia vs. other areas.	45
4.10	Distribution of values for feature 58A: obligatory possessive inflection.	46

4.11 Presence vs. absence of obligatorily possessed nouns in the Americas vs. other areas. 49

4.12 Presence vs. absence of obligatorily possessed nouns in Eurasia vs. elsewhere. . . 50

4.13 Distribution of values for feature 59A: possessive classification. 51

4.14 Two possessive classes vs. other types in Eurasia vs. other areas. 54

4.15 Distribution of values for feature 59A-r: possessive classification (reduced). 55

4.16 Presence vs. absence of possessive classification in Eurasia vs. elsewhere. 57

List of Figures

4.1	Geographical distribution of the unsampled data for locus of marking.	35
4.2	Geographical distribution of head-marking g-units.	35
4.3	Geographical distribution of dependent-marking g-units.	36
4.4	Geographical distribution of the unsampled data for the position of pronominal possessive affixes.	39
4.5	Geographical distribution of g-units with pronominal possessive prefixes.	40
4.6	Geographical distribution of g-units with pronominal possessive suffixes.	40
4.7	Geographical distribution of g-units with any pronominal possessive affixes.	44
4.8	Geographical distribution of g-units with no pronominal possessive affixes.	45
4.9	Geographical distribution of the unsampled data for obligatory possessive inflection.	47
4.10	Geographical distribution of g-units with obligatory possessive inflection.	48
4.11	Geographical distribution of g-units without obligatory possessive inflection.	49
4.12	Geographical distribution of the unsampled data for possessive classification.	52
4.13	Geographical distribution of g-units with no possessive classification.	52
4.14	Geographical distribution of g-units with two possessive classes.	53
4.15	Geographical distribution of g-units with three to five possessive classes.	53
4.16	Geographical distribution of g-units with more than five possessive classes.	54
4.17	Geographical distribution of g-units that have possessive classification.	56
4.18	Geographical distribution of g-units with no possessive classes.	56

Chapter 1

Introduction: Areal Typology

1.1 Goals and insights of linguistic typology

Linguistic typology has two overarching goals: (1) discover the extent of variation in linguistic structures in the languages of the world, and (2) infer patterns in that variation, or postulate constraints on them. Daniel (2011, p. 44) defines it thus:

Linguistic typology compares languages to learn how different languages are, to see how far these differences go, and to find out what generalizations can be made regarding cross-linguistic variations.

The generalizations made as part of linguistic typology can take different forms. Often the goals of typologists, especially early on, has been to identify universal structural features of language. Perhaps the most famous example of universals is Greenberg's (1963) list of apparent preferences that most or all languages have for certain word orders. What is notable about these universals is that they connect structural features—the behavior of a certain feature implies the behavior of a different feature. For example, Greenberg's second universal states, "In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes." When there is no strictly logical connection between the two features, but their behaviors do coincide, then there must be a structural property of the language itself that connects the two features. Such implicational universals thus reveal properties of language by means of comparing a large number of languages at the same time.

Unfortunately, Greenberg's sample of languages was probably not large enough (around 30 languages) or balanced enough to justify his generalizations about all the languages of the world. The problem of language sampling is a complex one that will be discussed in Chapter 3. Later researchers (for example Dryer, 1992, 2011c) have tested Greenberg's universals with larger

and more carefully selected samples. But Greenberg's work illustrates the goal of typology to find, through mass comparison, general properties of language that then need to be explained on a structural level.

Greenberg's universals, and universals that have been proposed since, are significant also in that they compare very specific features of language. Whereas earlier linguists classified languages under very broad terms such as **isolating** and **agglutinating** (terms which are still often used today), modern typology looks at more specific features of specific environments of languages. Bickel (2007, pp. 246–247) argues that “finer-grained variables” help solve the problem of cross-linguistic comparability: the more specific the variable, the less controversial it can be as to whether it is actually the same thing in different languages. Bickel gives the example of incorporation:

Instead of trying to decide whether structure S in language L is or is not incorporation, one codes S_L for a set of maximally fine-grained variables, just as large as to capture all that one knows about S_L (e.g., has generic reference: yes/no; prohibits permutation: yes/no; subcategorizes for a stem class: yes/no, must be adjacent to another stem: yes/no; satisfies argument slots: yes/no, triggers agreement: yes/no, or “NA” if the language has no agreement to begin with, etc.).

Not only do such variables ensure greater comparability across languages, but they also provide much richer data about the languages being studied. Instead of relying on a predetermined definition of incorporation and simply looking for its presence or absence, researchers should examine what looks as though it may be incorporation and then describe as much as possible. In this way, we can learn new aspects of how the proposed structure works with each studied language.

The World Atlas of Language Structures Online (Dryer & Haspelmath, 2011, henceforth WALSL) is a large database containing information on numerous features (currently 192) in all the major domains of language. Although these features may not be quite as specific as what Bickel describes, they do attest to how intricately languages can vary. The current study deals with the structure of possession, and WALSL contains data on eight features related to possession, four of which will be considered here (described in Chapter 2).

1.2 Areal influences on linguistic features

Structural universals are not the only kind of generalizations that can be made about language. As it has become possible to use larger and larger datasets for typological research, such as those in WALS, it has become evident that there are geographical influences at work in the preferences languages have for certain structures. In fact, Bickel (2007, p. 243) states that “hardly any typological variable, and only some combinations thereof, is evenly distributed in the world.” Even more important, he calls this uneven distribution “non-accidental,” meaning that there is likely to be some factor related to the location of the language that is influencing linguistic change. The kinds of forces that can drive change in a language are many, but Bickel (2011, p. 402) describe the two main principles that underlie them:

match There is some linguistic or cognitive reason for this structure to be preferred. Match factors constitute the linguistic universals described above, and explaining the reasons for them falls into the realm of linguistic theory, such as generative grammar.

spread Nearby languages have this feature, so situations of language contact increase the possibility of the language also taking on this feature.

If a significant number of unrelated languages in one geographic region seem to show a preference for a certain structure, while languages in other areas show different preferences, then it is likely that the structure has been diffused among the languages in that region, creating what is sometimes called a **Sprachbund** or **linguistic area**. Although Bickel (2007, p. 245) cautions against using such terms, since usually we are looking at individual features rather than broad categories, there is evidence of some areas in which multiple features are shared across unrelated languages—features which do not appear to be universally preferred but are instead specific to that particular area.

One example of a linguistic area that has been proposed is Mesoamerica. Campbell, Kaufman, and Smith-Stark (1986) find numerous linguistic traits shared among the languages of Mesoamerica that are not as frequent in other parts of the world, such as vowel harmony, fixed stress, the structure of nominal possession, relational nouns, and basic word order.

It is important to acknowledge that apparent groupings on a map alone are not enough to prove that a linguistic area exists. Bickel and Nichols (2005, p. 2; see also Bickel, 2007, pp. 243–244) emphasize that areal linguistics “must be grounded in a theory of population history, i.e. a theory of large-scale population and/or language movements—not on *visual* impressions.” In order to conclusively determine the existence of a linguistic area, one must have historical or anthropological evidence that language contact and borrowing has actually taken place. In this way, linguistic typology necessarily intersects with other social sciences in order to present an accurate view of what is going on with the distributional characteristics of linguistic features.

Even so, quantitative data that suggests areal behavior of features is a good starting point for deeper investigation into the actual reasons behind them. In that light, the goal of this study is to confirm that a linguistic area is likely with respect to certain features. The area in question is Eurasia, and the features to be investigated relate to possession.

1.3 The Eurasian macro-area

Eurasia has been identified as a linguistic macro-area by, among others, Nichols (1992), and later by Bickel and Nichols (2003, 2005) as part of the AUTOTYP project.¹ The AUTOTYP research program (Bickel, 2002) is similar to WALS in that it is a large database of languages coded for numerous variables. But there are some important differences between the two projects. Perhaps the most important difference is that categories are not defined before the data is gathered, but after. As new data from a language is entered into the database, its structures with respect to a particular variable are compared to the existing types to see if they fit into any of them. If they do not, then new types are defined based on the data, and the typology is automatically generated based on the data (hence the “auto” in AUTOTYP). Also, the project seeks to define variables as precisely as possible—the “finer-grained variables” that Bickel (2007) calls for—to ensure that variables are properly comparable across languages. These practices satisfy the demands of Haspelmath’s (2007) argument that there are no pre-defined categories in languages, and so linguists must instead focus on describing the structure of each language in as much detail as possible rather than trying to fit it into categories.

¹<http://www.spw.uzh.ch/autotyp/>

Note that the address has changed from what is given in many publications that mention the project.

Based on the data gathered in AUTOTYP, Bickel and Nichols have shown that there very well could be a Eurasian macro-area that stands apart typologically from the rest of the world. It is called a macro-area because it is apparently a combination of several areas, which Nichols (1992) calls spread zones. And although there is evidence of some typological enclaves within Eurasia that contrast with the general trends of the greater area (Bickel & Nichols, 2003), it can be considered as a whole because it “is characterized by a relatively uniform typological profile that contrasts with the rich structural diversity” in other areas. (Bickel, 2008, p. 228). Examples of features that define the Eurasian area are synthesis of verbal inflection and polypersonal agreement.

Although the AUTOTYP database is evidently larger and more complex, hence likely to be able to provide greater accuracy the problem, currently, is that the data is not easily obtained. Available publications reporting on the project give only summaries or highlights of the data and the exact steps taken to come to conclusions about Eurasia and other areas. Furthermore, persistent server errors make it difficult to obtain useful information on the project’s website. The goals of the current study share similarities with those of AUTOTYP, in that we are trying to determine the independence of Eurasia as a linguistic area with respect to particular variables. The primary difference is that this study uses publicly available data from WALS, as well as a transparent method that can be replicated, both for verifying these results and for testing other variables for which data is available. It should be unsurprising if the current method produces results similar to those obtained by AUTOTYP, since the data and method are probably quite similar (discussed in more detail in Chapter 3), but it is hoped that this study will cast more light on the issue by providing a technique that can be examined, repeated, verified, and extended to new applications. The data used here is readily available, and the sampling method applied to the data is clear and easy to repeat.

1.4 Possession as an areal feature

While Bickel and Nichols have tested a wide range of variables in the domains of phonology, morphology, and syntax, the current study focuses on a particular subdomain of language: the morphology of possession. Although from a European linguistic perspective the possessive construction may seem quite simple, the ways in which it can vary in languages across the world

are anything but simple, as we will see in Chapter 2. (In fact, the very notion of possession is notoriously difficult to define in a way that is cross-linguistically meaningful. This matter is also discussed in Chapter 2.) As mentioned above, the WALS database includes eight features related to possession. Some are syntactic rather than morphological, such as the order of the genitive and noun (Dryer, 2011a) and predicative possession (Stassen, 2011). The current study uses four particular features or variables that deal specifically with the morphology of possession. They are the following:

1. the locus of marking in the possessive noun phrase (Nichols & Bickel, 2011a)
2. the position of pronominal possessive affixes (Dryer, 2011b)
3. the presence or absence of obligatory possessive inflection (Bickel & Nichols, 2011)
4. the number of possessive classes (Nichols & Bickel, 2011d)

These features are examined in detail in Chapter 2, but the point we wish to emphasize here is they all have been noted to be unevenly distributed geographically. What is most intriguing is that the geographical skewings appear to be fairly similar across the different variables. For example, Nichols and Bickel (2011a), in describing the data for locus of marking, note that

Head-marked possessive NPs are common in the Americas and the Pacific (chiefly Melanesia) and infrequent elsewhere. Dependent-marked NPs have a roughly complementary distribution to this.

In speaking of the position of pronominal possessive affixes, Dryer (2011b) observes, The map shows what is perhaps the clearest apparent example in this atlas of an Old World–New World split in the distribution of the two types of possessive affixes: while possessive suffixes are the primary type in the Old World, possessive prefixes are primary in the New World.

The descriptions of the other two features share similar findings. The common thread is a roughly hemispherical division of linguistic structures—the Old World and the New World

(for lack of better terms) tend to have contrasting preferences when it comes to these possessive features. Furthermore, much of the skewing in the Old World appears to come from Eurasia specifically. In other words, Eurasian languages seem to contrast with the rest of the world.

However, the raw data from WALS cannot be taken at face value. In fact, immediately after the quotation above, Dryer advises wariness when identifying visual patterns on the maps. The exact reasons for taking caution in this regard are outlined in Chapter 3. But if, after some refinement of the data, these areal patterns continue to hold, the result would have interesting implications for the idea of Eurasia as a linguistic macro-area, especially since these features all surround one particular linguistic construction.

1.5 Overview of the thesis

The purpose of this study is to show whether Eurasia can be considered a linguistic area with respect to the features of possessive morphology listed above. In Chapter 2 we first tackle the question of how to define possession, and then we examine the details of each feature in turn.

The data for these features was taken from WALS. For reasons that are described in Chapter 3, the data cannot be used directly without some sort of controlled sampling. A technique of controlled sampling and its rationale is explained in that chapter. We also briefly touch on the statistical test that is used to determine the significance of the geographic distributions of the features, and then we explain how the data for this study was counted and compared.

In Chapter 4 we look at the results of the data, examining each feature in turn and how they are distributed among individual values before and after sampling. More important, we look at how these features are distributed geographically and test whether the languages of Eurasia (and other areas in some cases) behave independently of the rest of the world. We find that some features are significantly different while others are not.

Finally, in Chapter 5 we discuss the implications of these results. The notion of Eurasia as a linguistic macro-area is revisited. We explore some possible linguistic and historical reasons behind the unique behavior of Eurasian languages. We also evaluate the idea of possession considered as a unified typological feature. We conclude by acknowledging the limitations of this study and proposing directions for future research in this area.

Chapter 2

Features of possessive morphology

The notion of possession (and notions related to it) is expressed in a multitude of ways in the languages of the world, with a great deal of variation along multiple dimensions. This variation has been explored in a number of survey studies, most notably by Ultan (1978), Seiler (1983), Manzelli (1990), Croft (1990, pp. 28–39), Plank (1995), Rijkhoff (2002, pp. 86–91, 194–205), Koptjevskaja-Tamm (2003), and Dryer (2007, pp. 177–191). The studies vary widely in approach and scope, attesting to just how much variation is possible in an area as restricted as the possessive noun phrase.

As mentioned before, WALS includes data on several features related to the structure of possession. This study focuses on a few of those features, and in this chapter we will discuss the characteristics of each one. In Section 2.2, we will look at what is called the locus of marking, meaning where the marker showing possession is placed in relation to the constituents of the possessive noun phrase. A related feature is the use of suffixes versus prefixes to show possession, which will be discussed in Section 2.3. The concept of obligatory possessive inflection will be covered in Section 2.4, and in the final section we will look at possessive classification.

Before turning to the particular features, however, we must first investigate the crucial question of how exactly to define possession.

2.1 Defining possession

Although most speakers probably have an intuitive sense of what possession is (and is not) in their native languages, the task of clearly and unambiguously defining the phenomenon for the purposes of cross-linguistic comparison is far from simple (see, for example, Seiler, 1983, pp. 1–4 and Herslund & Baron, 2001, pp. 1–2 for expositions on the difficulties of defining it). There are a number of ways to approach this problem, but each way has its limitations.

The easiest definition is a semantic one, relying on the basic meaning of the word **possession** itself to denote a relationship of material ownership. Thus, *Mary's car*, *the neighbors' dog*, and *my phone* are prototypical examples of possessive noun phrases—one entity literally possesses the other. However, examples abound in English of noun phrases that are formally identical but could not be said to be expressing ownership:

- (1) a. John's arm
- b. my uncle
- c. your job
- d. the candidate's campaign
- e. the building's demolition
- f. today's topic

Each of these examples uses the same possessive marker 's for nominal possessors as the same set of possessive pronouns, yet each illustrates a different type of relationship (and there are certainly many more), none of which is actual ownership. Examples (1a) and (1b) are body-part and kinship relationships, respectively, which are commonly considered core meanings of possession. Koptjevskaja-Tamm (2002, 2003, p. 621) includes such relationships in her definition of possession along with legal ownership. Examples in many language grammars frequently express these types of relationships, especially when describing alienability distinctions (which we discuss in sections 2.4 and 2.5). But whether the other examples in (1) also express possession is less clear. In (1c), *your* could be considered to be possessing *job*, but certainly not in a material way. The possessor in (1d) is more like a subject in a nominalized clause, and the possessor in (1e) more like an object. The relationship in (1f) is difficult to classify and seems quite far from the core definition of possession; it is some kind of temporal relationship. All of these constructions are semantically quite widespread, yet they are all considered possessive constructions in English. If we are to rely on a semantic definition, it would need to be expanded.

Another option is to use a formal definition of possession. Seiler (1983, p. 4) notes that in terms of syntax, "POSSESSION is a relation between nominal and nominal, which is not mediated by a verb." This definition contrasts possession with other grammatical relations such as predication, which relates a verb to a nominal. It certainly captures all the examples in (1).

However, Ultan (1978, p. 13) contends that “a strictly or even primarily formal feature approach would lead to a fragmented and not particularly significant picture of how possessive systems are structured.” Relying solely on formal features would restrict our ability to describe and analyze the structural variation that is possible with possession. There appears to be a need for some semantic component. Croft (2003, p. 13), speaking of the issue of cross-linguistic comparability in general, argues that “the ultimate solution is a semantic one.”

Let us go back to the examples in (1). We have observed that none of the relationships expressed in these phrases denotes ownership, but each phrase does express some kind of relationship between the two entities. Without specifying what kind of relationship holds, we can say that possession necessarily encodes a semantic connection of some kind. This is the first component of our definition, making it broad enough to capture all the examples given. To avoid making it too broad, we should qualify the relationship by noting that it is *asymmetrical*: there are two entities in a possession relation, a **possessor** and a **possessum** (or **possessee**), and they are related in such a way that if the relation were reversed, the meaning would be fundamentally altered. Thus, *the neighbors’ dogs* and *the dog’s neighbors* are not equivalent.

A way of explaining this asymmetry is by viewing possession metaphorically as a type of location, as do Herslund and Baron (2001, p. 21) and Rijkhoff (2002, p. 175). Rijkhoff groups possessors with other dependent members of noun phrases denoting location, such as demonstratives, because a possessor delimits, at least metaphorically, where the possessum is located. Although the metaphorical extension is longer in some possessive relations than in others (it is harder to see in phrases like (1e), in which the formal possessor is the object of a nominalized verb), the important point is that possession restricts the set of possible referents of the possessum, as opposed to merely adding descriptive information as a modifier does. It is because of this restrictive function, Rijkhoff (p. 24) speculates, that possessor phrases as well as relative clauses appear more frequently as part of noun phrases than do adjectives. Herslund and Baron (2001, p. 21) also observe that the entities of a possession relation “receive their semantic interpretation in virtue of one another,” and their interpretation is not reciprocal, i.e. the relationship is asymmetrical.

Our definition of possession for this study, then, is *an asymmetrical relation between two nominal or pronominal entities, without a mediating verb*. We might add, for convenience, that

possession typically involves ownership of material objects, body-part relationships, kinship relations, or relations that are metaphorical extensions of these basic ones. Most of the examples given in this study express these typical kinds of relationships.

It is worth noting how the features discussed in this chapter fit into the larger context of possession as a whole. Possession could be said to fall into four broad categories. The first we will call adnominal, or in other words, the possessive noun phrase,¹ which is the focus here. It is indeed only to this type of possession that the definition developed above could apply. The other types present some challenges to the definition, but since they are mainly outside the scope of the present study, the definition works.

We mention the other three categories of possession briefly here. First there is predicative possession, which is commonly known as the ‘have’ construction, although Stassen’s (2011) study in WALs shows that ‘have’ is only one of several types of predicative possession. Although this is a fruitful area of study (see for example Stassen, 2009), it will not be dealt with here. Second, substantival possession is characterized by standalone words, usually pronouns, which express the concept of possession without an overt binary relationship, such as English *mine* and *yours*. Third, external possession is a relation between entities which are not in the same noun phrase, as in the sentence *I punched him in the teeth*; *him* and *teeth* are semantically related, but the structural connection is different from that of adnominal possession (see Payne & Barshi, 1999 for a collection of studies on specific languages having this feature). Each of these categories is a testament to the great complexity and variation exhibited by possession, but here we concentrate on features dealing with adnominal possession, or the possessive noun phrase proper.

We now turn to the specific features that will be examined in this study.

2.2 Locus of marking

The term **locus of marking** is attributed to Bickel and Nichols (2007; see also, Nichols & Bickel, 2011a, 2011b, 2011c), but the phenomenon itself was first called attention to as a typological variable by Nichols (1986). Essentially, this variable deals with where in the possessive noun phrase the possession morphology is placed, if any such morphology exists in the language. Af-

¹Manzelli (1990) applies the term “possessive adnominal modifiers” only to constructions with pronominal possessors, but we use it in a broader sense here to include noun phrases with both nominal and pronominal possessors.

fixes are found to be marking either the head of the phrase (the possessum) or the dependent (the possessor). English is an example of a **dependent-marking** language (refer again to (1)). In fact, Koptjevskaja-Tamm (2003, p. 626) has found that the vast majority of European languages mark the dependent of possessive noun phrases. Many languages in other parts of the world show this pattern as well. Examples are given below from Kannada (Dravidian, India) and Lezgian (Nakh-Daghestanian, Azerbaijan).

(2) huDugana haNaavannu kadiyabe:Da. Kannada (Sridhar, 1990, ex. 480)
 boy-POSS money-ACC steal-PROH
 ‘Don’t steal the boy’s money.’

(3) Mizafer.a-n k’wal-er Lezgian (Haspelmath, 1993b, p. 84)
 Mizafer-GEN house-PL
 ‘Mizafer’s house’

Hungarian is one of the only **head-marking** languages in Europe, but the pattern is much more prevalent in other parts of the world, as exemplified by Acoma (Keresan, New Mexico).

(4) az ember ház-a Hungarian
 the person house-3SG.POSS
 ‘the person’s house’

(5) s’adyúm’ə gâam’a Acoma (Miller, 1965, p. 177)
 1SG.brother 3SG.house
 ‘my brother’s house’

In (5), two possessive relationships are expressed: one between the speaker and ‘brother’, and the other between ‘brother’ and ‘house’. In both parts, the possessum is the word that receives the marking.

A subset of the dependent-marking languages in Koptjevskaja-Tamm’s (2003) survey also mark the head of the possessive noun phrase—the **double-marking** strategy. Nichols and Bickel (2011a) observe this structure to be rare, but instances of it are spread out all over the world. In Evenki (Altaic, Siberia and Northern China) the possessor is sometimes marked with a genitive suffix *-ngi*, as in (6a), although now it is more commonly used in the unmarked nominative, as in (6b). Possessive pronouns, as in (7), are formed from the genitive suffix. Nichols and Bickel (2011a) count Evenki as a double-marking language.

- (6) a. etyrken-ngi d'u-n Evenki (Nedjalkov, 1997, p. 82)
 old.man-POSS house-3SG.POSS
 'the old man's house'
- b. etyrken d'u-n
 old.man house-3SG.POSS
 'the old man's house'
- (7) minngi amin-mi
 my father-1SG.POSS
 'my father'

This head-dependent distinction is not as clear-cut when the language uses an independent word to show possession, rather than an affix or clitic. For example, a language may use an adposition, which is an optional variant in English but the only choice in Spanish, characteristic of many Romance languages.

- (8) the Queen of England
- (9) la familia del hombre Spanish
 the.FEM house of.the man
 'the man's family'

Dryer (2007, p. 179) classifies the use of adpositions as a type of dependent-marking, "since the adposition forms a constituent with the possessor." Adpositions are not the only type of independent possession word, however. Koptjevskaja-Tamm (2003, p. 671) cites "traditional analyses" which give the following tripartite distinction of associations, which corresponds to the locus of marking with bound forms: (a) dependent-associated forms, including prepositions and agreeing possessive "articles" (words that precede the possessor but show agreement features of the possessum); (b) head-associated forms, including linking pronouns (declined possessive pronouns that occur between the head and the dependent) and non-case-marking dependents; and (c) double-associated forms, including linking pronouns and case-marked dependents. In short, the possession word is often associated with one of the constituents of the phrase, even if not directly attached to it. Moru (Nilo-Saharan, South Sudan) is an example of a language that uses dependent-associated forms.

- (10) dri ts^wé ró Moru (Tucker & Bryan, 1966, p. 56)
 head tree of
 'the top of the tree'

It is possible for languages to use **no marking** at all to show possession. This strategy is also called **juxtaposition**, because in this structure two or more nouns are simply placed next to each other, and their consecutiveness implies the possession relationship.

- (11) teman ayah saya Indonesian
 friend father I
 'my father's friend'
- (12) Tomáása rooriy Yagua (Peba-Yagua, Peru; Dryer, 2007, p. 184)
 Tom house
 'Tom's house'

As shown in (11), nested possession relationships are possible within the same phrase in Indonesian. The relationships are simply evaluated in the order in which the nouns are juxtaposed, with the head of the phrase at the beginning.

Nichols and Bickel (2011a) also cite examples of marking strategies that do not fall cleanly into any of the above categories. Some languages use a clitic rather than an affix, the difference being that while an affix attaches to a specific member of the phrase (the head or the dependent), a clitic is placed in a particular position of the phrase, thus operating above the word level. The most well-known behavior of clitics is to gravitate toward the second (Wackernagel) position of the phrase, thus often attaching to the first word. Similarly, Dryer (2007, pp. 179–80) notes that possessive clitics usually appear immediately before the possessum. Such behavior is demonstrated in Chamorro (Austronesian, Guam):

- (13) a. i=lepblo=n estudiante Chamorro (Topping, 1973/1980, p. 208)
 ART=book=LINK student
 'the student's book'
- b. i=dankalo=n taotao
 ART=big=LINK man
 'the big man'

The *n* clitic in (13) serves a generic linking function, and so it can be used both for modifiers and for possession. Nichols and Bickel (2011a) point out that when the dependent is an attributive modifier, both attribute-noun and noun-attribute word orders are possible, but in possessive noun phrases the possessum must come first, as in (13a). In that phrase, the linking clitic attaches to the head, but in the inverted word order in (13b), the clitic still attaches to the same position.

Still, possession strategies like this one and a few others are rare, according to Nichols and Bickel's survey. Most languages are instances of structures that can be defined in terms of head or dependents.

2.3 Position of pronominal possessive affixes

The languages of the world vary greatly in terms of what exactly their possession markers mean. In English, for example, nominal possessors always take the 's ending to show a possessive relationship and nothing more. Hungarian, in contrast, marks the possessum with an affix that varies according to the person and number features of the possessor.²

- (14) a. a(z én) ház-am Hungarian
the (1SG.NOM) house-1SG.POSS
'my house'
- b. a (te) ház-ad
the (2SG.NOM) house-2SG.POSS
'your house'
- c. az ember ház-a
the person house-3SG.POSS.SG
'the person's house'
- d. a (mi) ház-unk
the (1PL.NOM) house-1PL.POSS
'our house'
- e. a (ti) ház-atok
the (2PL.NOM) house-2PL.POSS
'your(PL) house'
- f. a(z ő) ház-uk
the (3SG.NOM) house-3PL.POSS
'their house'
- g. az emberek ház-a
the people house-3SG.POSS
'the people's house'

²The optional pronouns are used solely for emphasis and are in the nominative case. The definite article *a/az* varies according to whether the next word begins with a consonant or a vowel.

The Hungarian possessive marking on the head noun agrees in person and number with the possessor.³ The type of marking exemplified in Hungarian is true **possessive** marking, while the marking used in English is called the **genitive**. The genitive is a grammatical case, meaning it shows a grammatical relation between two elements in an utterance.⁴ While it may sometimes vary according to features of the word to which it is attached, it does not reflect features of the other member of the phrase. Possessive marking, beyond showing a grammatical relation, shows features of the other item in the relation. Basically, possessive marking encodes more detailed information than does the genitive alone—morphological information that identifies the other member, such as person, number and even sometimes gender. Hungarian possessive marking, for example, shows the person and number of the possessor. Nichols and Bickel (2011a) have found that head-marked noun phrases most commonly show agreement features like these. When the marking is found on the head of the possessive noun phrase (as an affix), and it shows features of the dependent, it is often called a **pronominal possessive affix**, since it functions much like a pronoun in showing person, number, or gender features of the possessor, usually eliminating the need for an extra word when the possessor is pronominal. The Hungarian phrases in (14) show examples of pronominal possessive affixes.

Dryer's (2011b) survey in WALS deals with the exact position of pronominal possessive affixes—whether the language uses prefixes, suffixes, both, or neither. The main groups of interest include only head-marking languages in which the marking shows features of the possessor, because the survey is looking at affixes that act as pronominal possessors, as discussed above. Languages without such marking are put in the “no affixes” group, which comprises a sizable portion of the survey.

Hungarian uses possessive suffixes. Macushi (Cariban, Guyana) and Dumi (Sino-Tibetan, Nepal) are examples of languages that uses possessive prefixes:

³A notable exception to this agreement is when the possessor is third-person plural. If it is a noun, the marking on the head takes the singular form, as in (14g). If it is a pronoun, or not expressed at all, the plural possessive marker is used, as in (14f).

⁴Seiler (1983, p. 39) argues that, specifically, a verb is necessarily connected to a case-marked item: “Cases are means of expression that always contract some relation with the predicate or main verb. There is no exclusively adnominal case. This is true even for the genitive. Insofar as case forms contribute to the expression of POSSESSION, it is always by intermediacy of the verb.” It is not always clear how the verb can be connected; indeed, this statement seems to contradict Seiler's above-mentioned syntactic definition of possession. However, examples of languages exist in which the genitive is sometimes used for predicate relations and not for possession, such as Russian.

(15) u-pana' Macushi (Abbott, 1991, p. 86)
 1.POSS-ear
 'my ear'

(16) O:-ni ri:-t-a. Dumi (van Driem, 1993, p. 84)
 my-mind reel-NPT-23s
 'I'm dizzy.' (lit. 'My mind is reeling.')

Some languages use either prefix, depending on the circumstances. In Squamish (Salishan, British Columbia), whether the possessive marking is a prefix or a suffix depends on the person of the possessor (boldface in (17) indicates the possessive marking). Squamish can be said to have both prefixes and suffixes with neither primary.

- (17) a. ***ʔn**-snəx^oi'λ* 'my canoe' Squamish (Kuipers, 1967, p. 87)
 b. ***ʔə**-snəx^oi'λ* 'your (singular) canoe'
 c. *snəx^oi'λ-s* 'his canoe'
 d. *snəx^oi'λ-**čət*** 'our canoe'
 e. *ʔə-snəx^oi'λ-**ɨ**ap* 'your (plural) canoe'
 f. *snəx^oi'λ-s-**uit*** 'their canoe'

In Paumarí (Aruan, Brazil), many inalienable nouns receive both a prefix and a suffix to mark possession, depending on the person and gender of the possessor.

- (18) a. *o-gora-na* 'my house' Paumarí (Chapman & Derbyshire, 1991, p. 257)
 b. *i-gora-ni* 'your house'
 c. *gora-ni* 'her house'
 d. *gora-na* 'his house'
 e. *a-gora-na* 'our house'
 f. *ava-gora-na* 'your (pl) house'
 g. *va-gora-na* 'their house'

The prefixes in (18) give the most information as to the person and gender of the possessor, since the suffixes are mostly the same across the paradigm, but both prefix and suffix are necessary to express possession.

It is worth noting that pronominal (agreeing) possessive markers often bear a strong resemblance to marking in other areas of the grammar. In fact, Universal 1415 from the Universals Archive⁵ states, “IF heads of possessive constructions (=possesseees) agree with their possessors, THEN verbs agree with subjects.” It does not seem unreasonable to expect the actual shape of agreement morphology to be shared between these two types of relations. This prediction is borne out in Hungarian: the possessive ending for each person is quite similar to either the subjective or the objective present-tense verbal conjugation, as shown in Table 2.1.

	<i>lát</i> ‘see’		<i>arc</i> ‘face’
	SUBJ	OBJ	POSS
1SG	<i>lát-ok</i>	<i>lát-om</i>	<i>arc-om</i>
2SG	<i>lát-sz</i>	<i>lát-od</i>	<i>arc-od</i>
3SG	<i>lát</i>	<i>lát-ja</i>	<i>arc-a</i>
1PL	<i>lát-unk</i>	<i>lát-juk</i>	<i>arc-unk</i>
2PL	<i>lát-tok</i>	<i>lát-játok</i>	<i>arc-otok</i>
3PL	<i>lát-nak</i>	<i>lát-ják</i>	<i>arc-uk</i>

Table 2.1: Comparison of the two types of verb conjugation and possessive morphology in Hungarian.

Koptjevskaja-Tamm (2003, p. 691) notes a general trend in European languages which use possessive affixes like Hungarian: at least some adpositions in the language will be able to take the same affixes. Indeed, this pattern also holds in Hungarian. The postpositions are marked with possessive morphology to show that they are related to pronominal entities. Seiler (1983, p. 22) observes even more specifically that possessive pronouns are related in form to object rather than subject pronouns. This is true at least for English (compare *him* and *his* versus *he* and *his*, and note how *her* is used for both possessive and object pronouns). Manzelli (1990, p. 66) makes similar observations.

A possible explanation for these similarities, at least with verbal morphology, is offered by Ultan (1978, p. 29): Possessive noun phrases can be said to have a topic-comment structure similar to that of clauses, with the possessor being the topic and the possessum the comment. Since the predicate of a clause, the principal component of which is the verb, is considered the

⁵<http://typo.uni-konstanz.de/archive/intro/>

comment portion of an utterance, it would be unsurprising for the comments in both structures (clauses and noun phrases) to share morphology, i.e., subject agreement features on the verb and possessor agreement features on the possessum. For this reason, it is much more common for possessive pronouns (or possessive morphemes on the heads of noun phrases) to be related to the object pronouns rather than to subject pronouns (see also Seiler, 1983, p. 22).

2.4 Obligatory possessive inflection

One aspect of possession that has been the concern of a great deal of research (for example Seiler, 1983) is the concept of **alienability**. The features described in this and the following section both deal with alienability, in slightly different ways.

The alienability of an entity refers to the degree to which it is considered inherently connected to another entity. It is an apparently cultural concept that manifests itself in some languages, chiefly in possessive constructions. While many languages make no linguistic distinction with regards to alienability, there are many others in which the possessive structure changes depending on the alienability of the possessum.

Many languages (which are necessarily head-marking) have sets of nouns that cannot be used in an utterance without some form of possessive marking. In other words, they are **obligatorily possessed**. Conceptually, such nouns are inalienable, meaning they are connected to the possessor in such a way that they cannot be transferred. Alienable entities, on the other hand, are understood to be connected to their possessors by some sort of transaction such as purchasing, giving, stealing, and so on. The dividing line between the two groups does not always fall in the same place across languages. Typically, obligatorily possessed nouns include body-part and kinship terms (it is for this reason that Koptjevskaja-Tamm, 2003 includes these terms as essential in her overall definition of possession).

It is easy to confuse obligatorily possessed nouns with relational nouns, a concept described thus by Seiler (1983, p. 11):

A relational noun opens a position for another nominal in a way comparable to a verb that opens positions or places for arguments. Thus *father*, *head*, *name*, etc. are relational nouns in English in the sense that a “father” is always “someone’s father”, etc. Absolute nouns, like the English *water*, *rock*, etc. do not have this property.

However, Bickel and Nichols (2011) restrict the definition of obligatorily possessed nouns even further: they can appear only with possessive inflection, rather than merely prototypically so. In the words of Rijkhoff (2002, p. 87) “they are two-place, relational nouns”, and both places must be filled. In addition, Bickel and Nichols argue that this category is possible only in head-marking languages, which English is not, so Seiler’s examples do not work under this definition. Bickel and Nichols offer a rule of thumb for identifying obligatorily possessed nouns:

In practical dictionaries obligatorily possessed nouns are usually cited in one or another of the possessive forms. ... In scientific works they are often cited in stem form with a hyphen.

They give some examples from Amerindian languages that demonstrate this pattern:

- (19) a. -be’ ‘milk’ Navajo (Young & Morgan, 1987)
b. bi-be’ ‘her milk’
- (20) a. -jaza’ ‘horn’ Acoma (Miller, 1965)
b. záça’ ‘his horn’

The “nouns” in (19a) and (20a) are ungrammatical by themselves, without being attached to a possessor. However, Bickel and Nichols (2011) observe that many languages with this class of noun have word processes to let these words stand on their own. In Navajo, for example, such a word may be attached to a generic or indefinite possessor (e.g. -be’ ‘milk’ becomes *’a-be* ‘something’s milk, someone’s milk, some animal’s milk’), so that the obligatorily possessed noun may now stand on its own. But it still includes a possessor.

Obligatory possession illustrates the insight from Ultan (1978, p. 25) that “the degree of intimacy of the relationship is mirrored by the degree of linkage.” The inalienable nouns discussed in this section are inseparably connected to their possessors. While this pattern does not necessarily occur in all languages with alienability distinctions, it is interesting to see the morphosyntactic relationship reflecting the conceptual relationship in such an extreme way.

2.5 Possessive classification

Alienable and inalienable nouns are examples of possessive classes, in which each class has a different way of expressing a possessive relationship. The obligatorily possessed nouns we

is where it is important to remember that possessive classification is only a lexical property, because when multiple classes are present it is usually not clear what semantic features might be used to distinguish them. When a language has multiple possessive classes, they do not appear to differ by degrees of alienability, but rather they are just different ways of expressing possession for different sets of lexemes.

Languages with complex systems may indicate the different classes in complex ways. A striking example is found in Chichimeca-Jonaz (Oto-Manguenan, Mexico), which has at least twelve possessive classes “defined broadly as involving tone changes, prefix-like elements, various internal changes, and combinations of these as well as suppletions” (Nichols & Bickel, 2011d). Examples of the many classes are given in Table 2.2. The types of changes that occur are vastly different, and the changes can be so complex the word becomes hardly recognizable from one possessor to the next.

1SG	2SG	
<i>námen?</i>	<i>namén?</i>	‘face’
<i>suní</i>	<i>síni</i>	‘lip’
<i>kútún</i>	<i>utún</i>	‘neck’
<i>túmbi?ir</i>	<i>nímbi?ir</i>	‘tail’
<i>nahí</i>	<i>únho</i>	‘friend’
<i>tásóc?</i>	<i>kisóc</i>	‘belt’
<i>namá</i>	<i>éMə</i>	‘carrying rack’
<i>kúndí</i>	<i>kirí</i>	‘water’
<i>nambá</i>	<i>úngwa</i>	‘hat’
<i>kúmbo?</i>	<i>kibó?</i>	‘land’
<i>ka?á</i>	<i>kan?a</i>	‘hand’
<i>masú</i>	<i>uni?í</i>	‘wife’
<i>táta</i>	<i>úngwæ</i>	‘father’

Table 2.2: Examples of the possessive classes of Chichimeca-Jonaz (Lastra de Suárez, 1984, pp. 24–25).

Such complex systems appear quite exotic to speakers of European languages, but Nichols and Bickel (2011d) point out that these systems are comparable to the case declension and verb conjugation systems found in many European languages. Viewed in this light, it is not altogether

unexpected for a language to show so many different ways of encoding a grammatical relationship.

2.6 Conclusion

In this chapter we looked at several features dealing with the morphology of possessive constructions. Each of these features is included in WALS, with data for many languages of the world. In Chapter 4 we will investigate this data, using the technique described in Chapter 3, to find geographical patterns in the distributions of these features.

Chapter 3

Method

3.1 Problems with language sampling

As stated in Chapter 1, typological studies generally aim either to enumerate the great variation that is possible in linguistic structures across languages, or to find explainable patterns in the variation. In either case, a typological study needs to consider a large number of languages in order to make meaningful observations. The most accurate and revealing study would take every language into consideration, but it is obviously impossible to do so for a number of reasons. For one, an accurate picture of the possibilities of human language would require gathering data on every language that now exists, has ever existed, or ever will exist—a clearly unreasonable feat. But even when diachronic concerns are set aside,¹ we are never likely to have data on every currently spoken language. As with most statistical studies, a sample needs to be used.

Creating a language sample carries many of the same potential pitfalls as creating a sample for any kind of study, but some problems are unique to linguistic typology. The biggest hurdle is that of all the languages that are believed to be currently spoken in the world, only a fraction are documented. Lewis (2009) estimates that there are 6,909 living languages, but WALS contains data on only 2,650 languages. Even worse, most of the languages that are yet undocumented, “often spoken in isolated areas of the world and belonging to under-investigated language groups, and which potentially harbor unique features, are on the brink of extinction” (Bakker, 2011, p. 100). There may be linguistic features we will never know about, and which could make significant differences in the conclusions drawn by typologists, simply because languages die out before they can be documented.

¹Perkins (2001, p. 423) assures that although “we have no information concerning most of the languages of the past, and none concerning languages that do not yet exist,” it is reasonable to assume “that such languages resemble languages for which records and evidence now exist.” There is, of course, no way of knowing if this is accurate, but we have no choice but to proceed on that assumption.

Furthermore, not all available documentation is equally valuable. Field linguists do not always go looking for the same features in the languages they research, and not every description has the same level of detail. As an illustration, although WALS contains data on a total of 2,650 languages, Nichols and Bickel (2011a) in surveying the locus of marking in possessive noun phrases found relevant data in only 236 languages—less than 10 percent of all the WALS languages. Many features in WALS have even less data. This problem stems from the fact that there is no standardized method of documenting languages, so from one description to the next there may be vast differences in the kinds of information provided.² Some language descriptions, particularly older ones, have the additional problem of bringing strong cultural or linguistic bias, thus calling into question the reliability of the information. These difficulties are at best extremely difficult to overcome, so typological studies must simply work with the best samples that can be obtained.

Rijkhoff, Bakker, Hengevald, and Kahrel (1993) and Bakker (2011) stress that there are two types of language samples, and the type of sample that should be obtained depends on the type of typological question that one is trying to answer. These two types correspond with the two overall goals of linguistic typology described above and in Chapter 1. The first goal of revealing all the possible variation of a certain structure calls for what is called a **variety sample**: “If ... one tries to account for all possible realizations of a certain meaning, like definiteness or relative clause, then the sample should display the greatest possible diversity” (Rijkhoff et al., 1993, p. 171). WALS is essentially a set of this kind of sample, collecting as much information as possibly about as many languages as possible. However, simply using every possible source of data available may result in a great deal of effort for little gain, and so Rijkhoff et al. (1993) developed a method for determining the diversity of a sample in order to create an optimally diverse sample for variety studies.

The other type of sample described by Bakker is the **probability sample**, which is concerned with helping to determine the probability of a language having a certain feature (or form of a feature). This kind of sample has very different goals and thus should be constructed differ-

²The Lingua Descriptive Studies Questionnaire (Comrie & Smith, 1977) attempts to mitigate this problem by outlining a systematic, thorough, and efficient way for language documenters to gather information that will be useful for typologists. Unfortunately, only a limited number of languages have been documented using this questionnaire. (The questionnaire itself can be viewed at <http://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaire/linguaQ.php>.)

ently. Size is one major difference: whereas a variety sample needs to be quite large, a probability sample tends to be small (defined by Bakker as “typically between 50 and 200 languages”). The reason it is smaller is that each language in the sample needs to be independent. Perkins (1989, p. 299) defines independence as meaning that “one cannot predict with better than chance odds the probability of any particular type of item being chosen based on one’s knowledge of another of its properties”. As applied to linguistic typology, “Independence means that the values of a variable are not predictable with greater than chance odds given the values of some other variable measurable in the sample” (Perkins, 2001, p. 427). For example, one should not be able to predict that a language has a certain feature based on its location. The larger the sample, Perkins warns, the less chance of independence, with the result of an increased possibility of systematic errors.

Problems arise when these two types of samples are conflated. When a sample that is optimized for diversity is used to make statistical inferences, spurious conclusions may result because of biases in the sample. Rijkhoff et al. (1993, p. 172) list five major types of bias that may occur in a sample: genetic, geographic, typological, cultural, and bibliographic. They argue that the most dangerous of these biases is genetic—how languages are related to each other by common descent—because it often leads to other kinds of bias.

Dryer (2009) warns that samples drawn from WALS data are especially prone to genetic bias. As mentioned before, the database draws as much information as possible for as many languages as possible, with essentially no filter. The result is a rich set of data on many features, which is useful for showing diversity, but it is problematic for statistical tests because some language groups are disproportionately represented. This genetic bias stems from the bibliographic bias discussed above; some languages are simply better known to linguists, and hence more thoroughly documented. The most prominent example of this bias is the Indo-European family, whose member languages are represented in large numbers on many of the WALS maps. For many features, Indo-European languages are mostly of the same type, so a large group of languages may be overwhelming languages of different types not because that type is preferable but because they all descend from a common ancestor. Because there are so many Indo-European languages recorded in WALS, it may appear that a certain type is more common in a certain area or overall, when in reality those languages are in the minority.

Genetic bias in language samples is clearly a serious problem that must be addressed in typological studies. Below we discuss a method that has been developed to try to overcome this problem. We will use this method to control for genetic bias as we examine morphological features of possession.

3.2 A controlled sampling method

3.2.1 Dryer's genera-counting method

A number of different methods of language sampling have been developed to maximize the independence of the languages (Bell, 1978; Perkins, 1989). The present study focuses on Dryer's (1989) method (refined by Bickel, 2008, discussed below), which he demonstrates in several studies using word order as an example (1992, 1997, 2000, 2003, 2011c). The idea driving this method is that rather than counting languages as individual cases of certain type, one counts the genera containing languages of that type. A genus is defined as a group "roughly comparable to the subfamilies of Indo-European, like Germanic and Romance" (p. 267). The languages in WALS are divided into such genera (some languages have additional levels of grouping, but all languages in the database are grouped on at least three levels: family, genus, and language). Counting at the genus level, Dryer argues, eliminates most of the genetic bias in language samples, since languages of the same genus tend to be similar typologically. Therefore, if all of the languages of a genus are of type A, then one unit of type A is counted. But if the genus contains languages both of type A and of type B, then one unit of each type is counted.

The genera are organized into several large geographic areas, and the counts are totaled for each area separately. These areas are assumed to be linguistically independent of each other, the idea being that if the same type or correlation of types is preferred in each independent area, then the tendency can be assumed to be universal. Dryer (1989) initially used five areas but later (1992) expanded the list to six: Africa, Eurasia, Southeast Asia & Oceania, Australia–New Guinea, North America, and South America. Table 3.1 illustrates the results of this method when comparing the order of object and verb to the order of noun and relative clause (Dryer, 1992, p. 86). Based on these results, we can see, for example, that languages with VO order strongly prefer NRel order.

	Africa	Eurasia	SEAsia&Oc	Aus-NG	NAmer	SAmer	Total
OV&RelN	5	11	2	2	3	3	16
OV&NRel	9	5	2	6	12	3	37
VO&RelN	0	0	1	0	0	0	1
VO&NRel	21	8	12	3	11	5	60

Table 3.1: Example of Dryer’s genera-counting method, comparing verb-object order and noun-relative order.

The advantage of dividing up the genera into these areas is that it accounts for another type of bias described by Rijkhoff et al. (1993): geographic bias. An area with a large number of reported languages will have no effect on an area with a smaller number of reported languages; the proportions of each area are considered individually. If the total numbers of each type were taken as a whole, highly skewed proportions could result. Furthermore, it is not always clear whether the individual languages in a given area are typologically independent, even if they are genetically independent, since borrowing or diffusion may have occurred (as discussed in Perkins, 1989). Dividing up the languages into areas believed to be independent and counting them separate reduces the effects of diffusion. Thus, Dryer’s method provides a solution for both genetic and geographic bias.

3.2.2 Bickel’s g-sample method

Dryer’s method is not without its weaknesses, however. Bickel (2008, pp. 224–226) identifies a few problems that are still present in this technique despite its advantages. One issue is what Bickel calls non-discreteness, or the fact that Dryer’s sampling “is an all-or-nothing issue.” The problem is that when a genus has more than one value of a variable in its member languages, the sampling does not take into account the actual distribution of those values; it simply provides one unit per value present in the genus. To use locus of marking as an example, if a genus had 6 head-marking languages and 3 dependent-marking languages, we would count each type once under Dryer’s (1989) method. However, Bickel argues, the distribution matters. Sometimes it is significantly skewed toward a certain value, in which case the majority value should be represented only once, since “it is likely (though by no means necessary!) that the distribution is induced by shared retention, innovation or family-bound drift.” On the other hand, sometimes

the distribution is not significantly skewed—i.e. it is no more than chance diversity. In that case, Bickel argues, “each language can be included without any reasonable risk of a genealogical inflation effect.” So in our 6-3 genus, if the skewing toward head-marking is determined to be significant, then that type should be counted once, but if it is not significant, then each individual head-marking language should be counted.

Another problem identified by Bickel is that the original method looks at a predetermined taxonomic level, the genus, but there may be skewing at other levels that would cause it. For example, if every language in a genus has one particular value of a variable, (e.g. every language in the genus is head-marking), it would be counted once, but it is possible that every other genus in the same family is skewed the same way. In that case, it is highly likely that all the genera inherited that property at the family level, so the value should be counted only once (e.g. head marking should be counted just once for the entire family).

Addressing these problems requires a more rigorous re-working of the method, since it means checking at all levels of the available taxonomy and testing for statistical skewing. Bickel has developed an algorithm called “g-sampling,” using the statistical package R, to automate the process.³ The input for the script is a table (such as can be obtained from WALS) that includes each language to be sampled along with each of its taxonomic levels (e.g. genus and family). Each language in the table is also coded for a response variable and, optionally, predictor variables. For example, if one were testing for a correlation between two typological features, like Dryer’s example shown in Table 3.1 on the previous page, the variables would be those two features—in the case of Dryer, the predictor would be verb-object order and the response would be noun-relative order. On the other hand, a predictor variable may be a geographical area, if one is testing for areal effects on a typological variable.

The algorithm organizes the languages into their genetic groups and then tests each group for skewing, starting at the top and going down. The result is a list of genealogical units (which Bickel calls “g-units”) that are each labeled with one of the following distribution types:

singleton If the highest-level taxon has only one member language, then it will be labeled as a singleton. Also, if a group is determined not to have significant skewing toward a value, then the individual members of that group will be labeled as singletons.

³<http://www.r-project.org>. Bickel’s script can be downloaded at <http://www.spw.uzh.ch/software>.

skewed (absolute) If every member of a taxon has the same value, then that taxon will be included as a single g-unit and marked as absolutely skewed.

skewed (trend) If the group is found to have significant skewing toward one value, then the majority value is counted as one g-unit and marked as having a trend.

deviate In a group that is skewed with a trend toward a certain value, the taxa that have different values will be marked as deviates, each deviating value counting as one g-unit.

This method further reduces the chance of genealogical bias. Values are represented based on their distribution, not simply based on whether they are present or not. As a result, large homogenous groups are reduced to a single unit, ensuring that they do not unjustifiably skew the results. For example, in many features in WALS, the Indo-European family is often reduced to a single point, because most or all of its languages have the same value for the feature. These g-units can then be counted and totaled in tables like Table 3.1. They can be assumed to be independent because of the reduction that has taken place.

3.3 Statistical tests

Related to the problem of language sampling is the question of what kinds of statistical tests are appropriate for typological studies. Although statistical techniques are not always necessary for valuable typological studies, according to Perkins (2001, p. 419), they usually are required if the researcher wants to make inferences from the data, as we do here. It is important to use the right tests; both Perkins (2001) and Cysouw (2005) caution that using the wrong types of tests can lead typologists to make erroneous conclusions.

One way that statistics are done inappropriately in typological studies is when parametric tests are used. Many common statistical tests are parametric, which means they assume that the sample is randomly selected. Janssen, Bickel, and Zúñiga (2006, p. 420) point out that random sampling is usually not feasible for linguistic typology, since the data is so limited as discussed above, and also because many language families appear to consist of a single language or very few languages, making it very difficult or impossible to select languages from each family in a truly random fashion. We have discussed one way of dealing with the sampling problem, but because the sample is not random, it is inappropriate to use tests that assume a random sample.

Non-parametric tests are preferable in this situation. For categorical variables, such as the features we are dealing with in this study, the Pearson chi-square test is commonly used to determine whether the distribution of cases (e.g. languages) among the different values is significantly different from a random distribution. However, Janssen et al. (2006, pp. 423–425) note that the chi-squared test requires a large amount of data so that expected values (the random distribution to be tested against) can be evenly distributed over the contingency table. But the required amount of data (for which there is no actual agreed-upon rule) is usually not available in typological studies. Furthermore, “empty cells and cells with small values are particularly interesting. They suggest heavy biases in the data, and yet these tables are intrinsically hard to test with the Pearson chi-square test” (p. 425). For this reason, a distribution-free non-parametric test is desirable, and Janssen et al. recommend Fisher’s exact test. Fisher’s exact is used for contingency tables like those used for chi-square tests, but it can accept small values, as well as highly uneven distributions, in the table. The present study will use Fisher’s exact test to examine the distributions of features.

3.4 The present study

The goal of the present study is to determine whether there are any areal differences between Eurasia and the rest of the world specifically, or the Old World and the New World generally, with respect to four features of possessive morphology: the locus of marking, the position of pronominal possessive affixes, the presence or absence of obligatory possessive inflection, and the number of possessive classes. The data for each of these features was obtained from WALS.⁴ One group of languages was removed from the database—the “other” family, which is really just the set of all pidgins, creoles, and sign languages in the database and has no genetic basis.

⁴The WALS data is available at <http://wals.info/export> in the form of one master table. For this study, individual tables for each feature was extracted from the master table, so that they contained only those languages that are coded for the particular feature.

The input tables for the g-sampling algorithm contained each relevant language coded with its genus and family.⁵ The predictor variable for each language was the value of the feature given in WALS. These tables were then run through the g-sampling script.⁶

After the g-sampling, each g-unit was given latitude and longitude coordinates. Each g-unit at the language level was given the coordinates listed for that language in WALS. For every g-unit at the family and genus level, the latitude and longitude were calculated by taking the mean of all the latitudes and longitudes (respectively) of its individual member languages.⁷ Using these assigned coordinates, the g-units were then plotted on a map of the world using Gnuplot.⁸ On these maps, the g-units were counted and grouped according to area (Africa, Americas, Eurasia, and the Pacific) and their value for the variable in question. The totals of these counts were put into contingency tables. For some of the features, depending on the claims made by the authors of the original WALS article on that particular feature, contingency tables were created to compare the Old World (Africa and Eurasia) with the New World (the Americas and the Pacific). In all cases, contingency tables were created for to compare Eurasia with the rest of the world. These tables were then statistically tested for independence using Fisher's exact test. If these tests yielded a *p*-value of less than 0.05, then the distribution was judged to be significantly different from random distribution, and thus there are apparent areal effects on the feature in question.

We look at the results of these tests in Chapter 4.

⁵As mentioned above, some languages listed in WALS have more than three levels of genetic grouping, but these additional levels were ignored, because the g-sampling script leaves out any levels that are not shared by all the languages in the table.

⁶The data files that were used for input in the g-sampling script, as well as the output files with latitude and longitude coordinates added, are available for download at <http://linguistics.byu.edu/thesisdata/garrettnay-possession.zip>.

⁷The average latitudes and longitudes for genera and families were calculated from the coordinates of every member language listed in WALS, not just the languages that are coded for the variable being looked at. This method is admittedly imprecise; it is discussed further in Chapter 5.

⁸<http://www.gnuplot.info>. The data file of coordinates for producing the world map in Gnuplot was obtained from <http://www.gnuplotting.org/plotting-the-world-revisited/>. The resulting maps are shown in Chapter 4.

Chapter 4

Results

In this chapter we discuss the results of the process outlined in Section 3.4. The four features to be discussed are locus of marking in the possessive noun phrase, position of pronominal possessive affixes, obligatory possessive inflection, and possessive classification. These variables are represented in WALS as features 24A, 57A, 58A, and 59A respectively.¹

Although the focus here is on the g-sampled data, summary information about the original WALS data is provided here for comparison. For purposes of space, the maps based on the unsampled data (Figures 4.1, 4.4, 4.9, and 4.12) display the datapoints for all values simultaneously. The result is an admittedly dense picture, but it should be adequate for general impressions and comparisons. The reader wishing to see more readable maps from the original data is referred to the locations listed in footnote 1. In contrast, the maps using g-sampled data are separated into individual values for each feature.

When the text refers to specific values from contingency tables, those values are underlined in the tables for convenience. The *p*-values for the contingency tables are obtained by Fisher's exact test.

4.1 Locus of marking

The unsampled data for feature 24A in WALS contains 236 languages with the distribution shown in Table 4.1.

¹The WALS data for each respective feature can be found at the following locations:

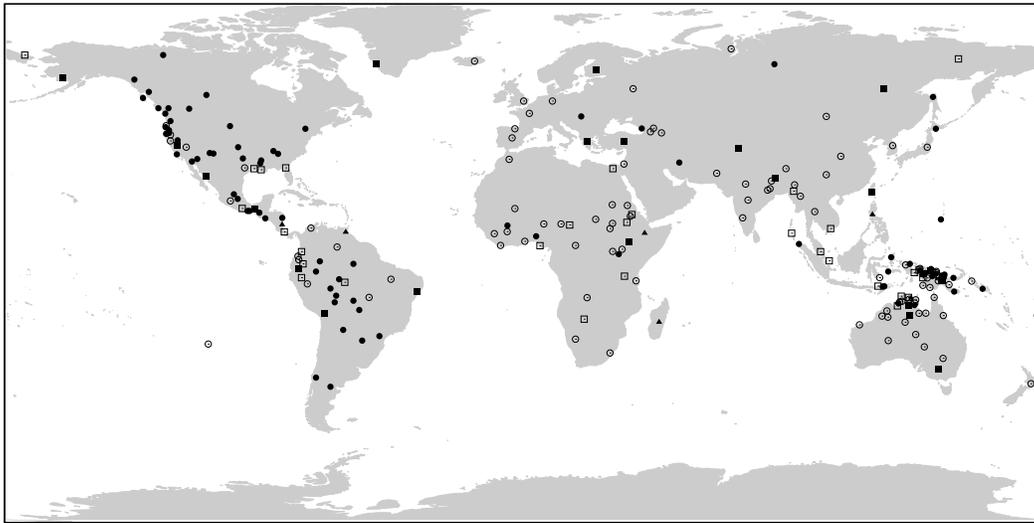
- <http://wals.info/feature/24A>
- <http://wals.info/feature/57A>
- <http://wals.info/feature/58A>
- <http://wals.info/feature/59A>

	Unsampled		G-sampled	
Head marking	78	33.05%	74	39.95%
Dependent marking	98	41.52%	57	30.00%
Double marking	22	9.32%	21	11.05%
No marking	32	13.56%	32	16.84%
Other	6	2.54%	6	3.16%
Total	236	100.00%	190	100.00%

Table 4.1: Distribution of values for feature 24A: locus of marking in the possessive noun phrase.

The two main values for this variable are head marking and dependent marking. Nichols and Bickel (2011a) note, and the WALS map shows, that head-marking languages are more common in the Americas and the Pacific, a general area that we will call the New World, than elsewhere. Dependent-marking languages are more common in Africa and Eurasia, which we will collectively call the Old World. In New Guinea, the types overlap. These observations are based on the geographical distribution of the unsampled data, shown in Figure 4.1 on the next page.

G-sampling this data resulted in 190 g-units, distributed as shown in the second part of Table 4.1. The difference between the sampled and unsampled data with respect to every value is not significant ($p = 0.184$). However, when only head marking and dependent marking are considered, the difference is significant ($p < 0.05$). In terms of independent g-units, dependent marking is revealed to be less common than head marking, contrary to the unsampled data. However, not much stock should be placed in these global differences, since they may be affected by geographical bias. The geographical distribution of head-marking g-units is displayed in Figure 4.2 on the following page, and the distribution of dependent-marking g-units is displayed in Figure 4.3.



- Head marking •
- Dependent marking ◊
- Double marking ■
- No marking □
- Other ▲

Figure 4.1: Geographical distribution of the unsampled data for locus of marking.

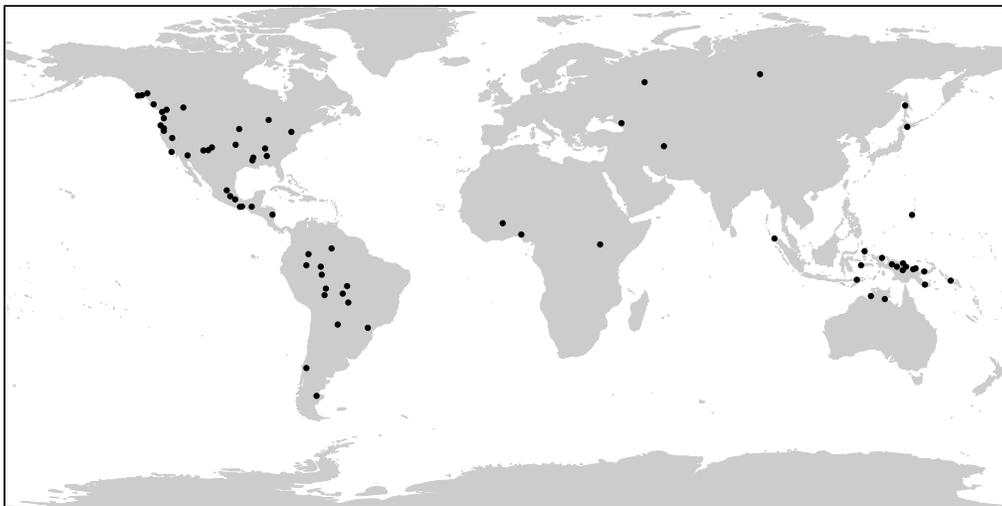


Figure 4.2: Geographical distribution of head-marking g-units.

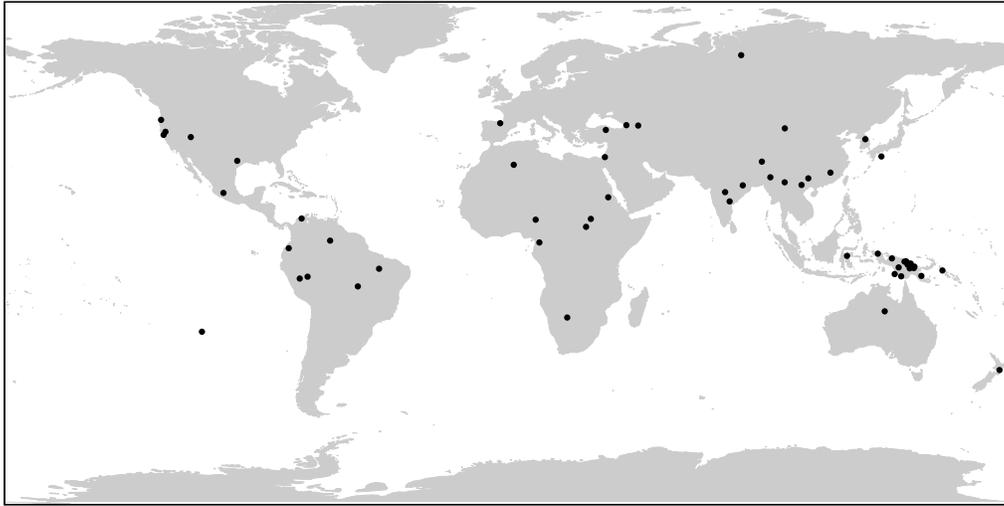


Figure 4.3: Geographical distribution of dependent-marking g-units.

Based solely on visual inspection, the hypothesis that head marking is more common in the Americas and the Pacific appears justified. Table 4.2a on the next page shows the counts of head-marking types versus all the other types in the New World versus the Old World. The distribution is highly skewed ($p < 0.005$), showing that in the Old World (namely, Eurasia and Africa), languages are far less likely to be head-marking than other types, with a column percentage of 16.39. Languages in the New World, on the other hand, are equally likely to be head-marking as any other type, the column percentage being 49.61.

When the dependent-marking type is singled out, the opposite situation appears to hold, as shown in Table 4.2b. The proportion of dependent-marking g-units in the New World is about one in four (24.03 percent), which is quite close to the total worldwide distribution of 30 percent. In the Old World, however, the division between dependent-marking and other types is nearly half-and-half (42.62 to 57.38 percent). This distribution is also significant ($p < 0.05$).

	(a)			(b)			
	NW	OW	Total	NW	OW	Total	
Head-marking	64	10	74	Dep-marking	31	26	57
<i>% of column</i>	<u>49.61</u>	<u>16.39</u>	<u>38.95</u>	<i>% of column</i>	<u>24.03</u>	<u>42.62</u>	<u>30.00</u>
<i>% of row</i>	86.49	13.51	100.00	<i>% of row</i>	54.39	45.61	100.00
Other	65	51	116	Other	98	35	133
<i>% of column</i>	50.39	83.61	61.05	<i>% of column</i>	75.97	<u>57.38</u>	70.00
<i>% of row</i>	56.03	43.97	100.00	<i>% of row</i>	73.68	26.32	100.00
Total	129	61	190	Total	129	61	190
<i>% of column</i>	100.00	100.00	100.00	<i>% of column</i>	100.00	100.00	100.00
<i>% of row</i>	67.89	32.11	100.00	<i>% of row</i>	67.89	32.11	100.00

Table 4.2: Head-marking vs. other types (a) and dependent-marking vs. other types (b) in the New World vs. the Old World.

When Eurasia is compared separately with the rest of the world, the proportions are nearly the same. Table 4.3a on the following page shows the distribution of head-marking g-units in Eurasia versus elsewhere, and Table 4.3b shows the distribution of dependent-marking g-units. Fisher’s exact test shows both of these distributions to be significant ($p < 0.005$ and $p < 0.05$, respectively). Note how in Table 4.3a, head-marking g-units make up 17.07 percent of total units in Eurasia, while in Table 4.2a they make up 16.39 percent of units in Eurasia and Africa combined. Similarly, 43.90 percent of g-units in Eurasia are dependent-marking, while 42.62 percent of g-units and Eurasia and Africa are dependent-marking. The data suggests that Eurasia is considerably skewed compared with the rest of the world when it comes to the locus of marking. Languages in Eurasia are highly likely to be dependent-marking, and highly unlikely to be head-marking.

	(a)			(b)			
	Eurasia	Other	Total	Eurasia	Other	Total	
Head-marking	7	67	74	Dep-marking	18	39	57
<i>% of column</i>	<i>17.07</i>	<i>44.97</i>	<i>38.95</i>	<i>% of column</i>	<i>43.90</i>	<i>26.17</i>	<i>30.00</i>
<i>% of row</i>	<i>9.46</i>	<i>90.54</i>	<i>100.00</i>	<i>% of row</i>	<i>31.58</i>	<i>68.42</i>	<i>100.00</i>
Other	34	82	116	Other	23	110	133
<i>% of column</i>	<i>82.93</i>	<i>55.03</i>	<i>61.05</i>	<i>% of column</i>	<i>56.10</i>	<i>73.83</i>	<i>70.00</i>
<i>% of row</i>	<i>29.31</i>	<i>70.69</i>	<i>100.00</i>	<i>% of row</i>	<i>17.29</i>	<i>82.71</i>	<i>100.00</i>
Total	41	149	190	Total	41	149	190
<i>% of column</i>	<i>100.00</i>	<i>100.00</i>	<i>100.00</i>	<i>% of column</i>	<i>100.00</i>	<i>100.00</i>	<i>100.00</i>
<i>% of row</i>	<i>21.58</i>	<i>78.42</i>	<i>100.00</i>	<i>% of row</i>	<i>21.58</i>	<i>78.42</i>	<i>100.00</i>

Table 4.3: Head-marking vs. other types (a) and dependent-marking vs. other types (b) in Eurasia vs. other areas.

The largest language families (based on how many languages have available data for the feature) with skewing toward a single value are Australian (27 languages), Niger-Congo (12 languages), and Indo-European and Nilo-Saharan (both 11 languages). All of these families are strongly skewed toward dependent marking. Among the Indo-European languages, only two exceptions are reported: Iranian (head-marking) and Modern Greek (double-marking).

4.2 Position of pronominal possessive affixes

The amount of raw data for WALS feature 57A is quite large (as it is for every feature surveyed by Dryer), consisting of 902 languages. The breakdown of each value for the feature is shown in Table 4.4.

	Unsampled		G-sampled	
Prefixes only	255	28.27%	108	33.54%
Suffixes only	355	39.36%	64	19.88%
Prefixes and suffixes	32	3.55%	29	8.07%
No affixes	260	28.92%	121	37.58%
Total	902	100.00%	322	100.00%

Table 4.4: Distribution of values for feature 57A: position of pronominal possessive affixes.

After the g-sampling algorithm, the data is pared down considerably, down to 322 g-units. Additionally, the new distribution among the values, shown also in Table 4.4, is significantly different ($p < 0.005$) from that of the unsampled data, the most notable difference being that g-units with only prefixes apparently now outnumber g-units with only suffixes. However, geographical factors must be taken into consideration. Figure 4.4 shows the map based on the unsampled data.

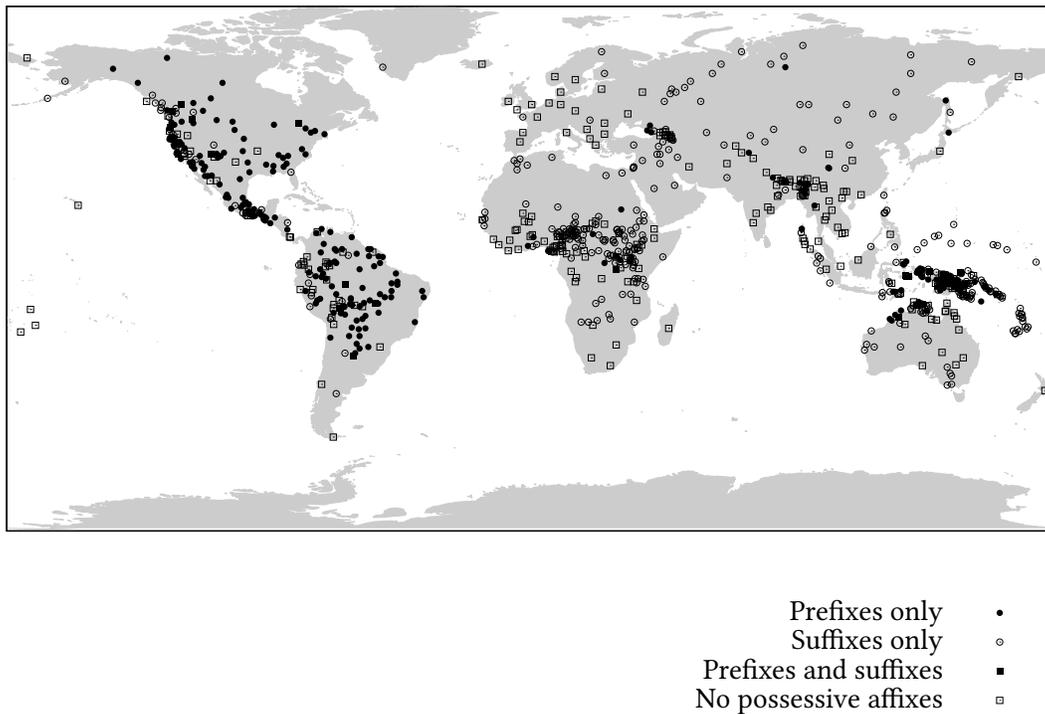


Figure 4.4: Geographical distribution of the unsampled data for the position of pronominal possessive affixes.

Dryer's (2011b) observation is that possessive prefixes are found primarily in the New World, and possessive suffixes are found primarily in the Old World. This situation does appear to be the case based on the first map. At first glance at the maps from the g-sampled data, however, it seems that the ratio of prefixes to suffixes is similar in both regions. The distribution of g-units with prefixes only is shown in Figure 4.5 on the next page, and the distribution of suffixes only in Figure 4.6.

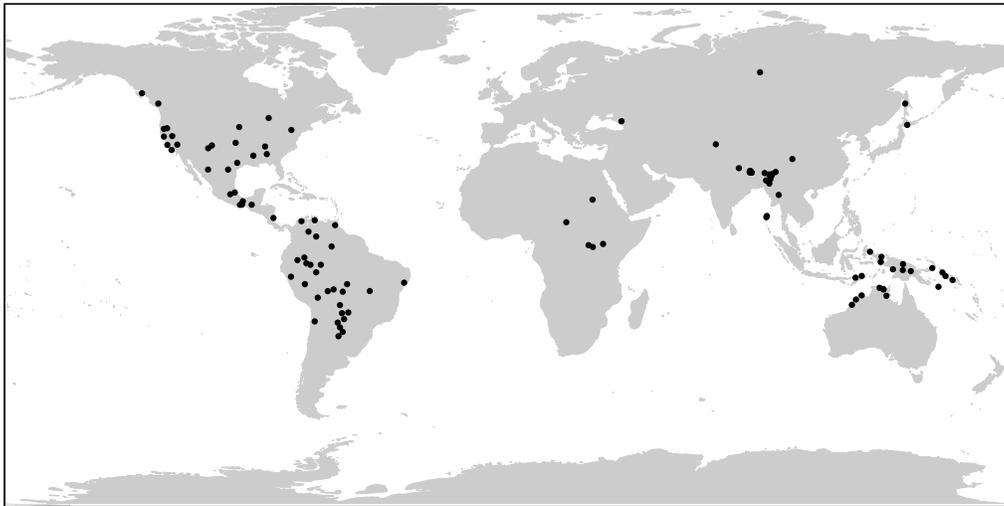


Figure 4.5: Geographical distribution of g-units with pronominal possessive prefixes.

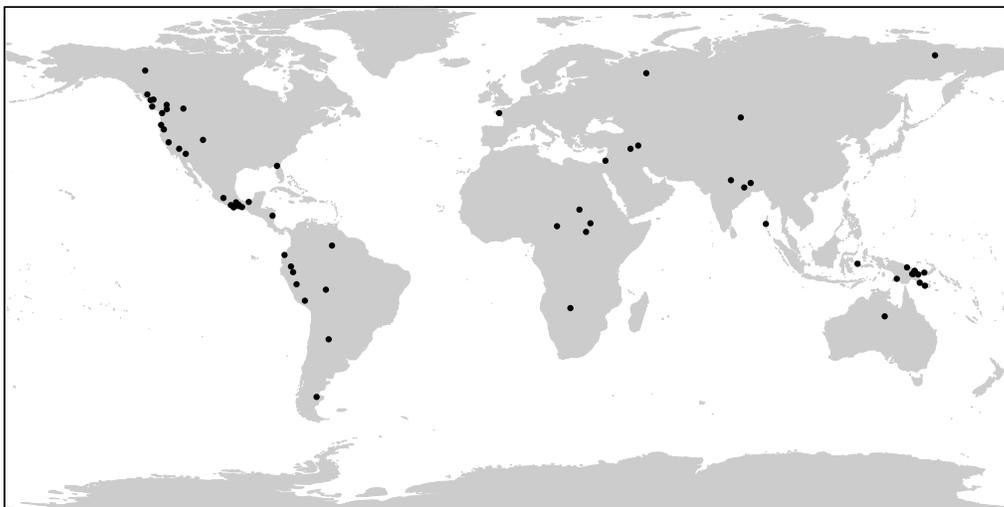


Figure 4.6: Geographical distribution of g-units with pronominal possessive suffixes.

The actual counts support this observed similarity between the regions. In Table 4.5a on the following page we see that there is a negligible difference ($p \approx 0.5$) between the two areas

with respect to the proportion of g-units with prefixes only; approximately one-third of g-units in both areas use possessive prefixes. In a similar fashion, there is no significant difference between the two areas in the proportion of suffix-using g-units, as shown in Table 4.5b ($p \approx 1$). About one-fifth of g-units in both areas use suffixes.

	(a)			(b)			
	NW	OW	Total	NW	OW	Total	
Prefixes only	79	29	108	Suffixes only	48	16	64
<i>% of column</i>	<u>32.51</u>	<u>36.71</u>	33.54	<i>% of column</i>	<u>19.75</u>	<u>20.25</u>	19.88
<i>% of row</i>	73.15	26.85	100.00	<i>% of row</i>	75.00	25.00	100.00
Other	164	50	214	Other	195	63	258
<i>% of column</i>	67.49	63.29	66.46	<i>% of column</i>	80.25	79.75	80.12
<i>% of row</i>	76.64	23.36	100.00	<i>% of row</i>	75.58	24.42	100.00
Total	243	79	322	Total	243	79	322
<i>% of column</i>	100.00	100.00	100.00	<i>% of column</i>	100.00	100.00	100.00
<i>% of row</i>	75.47	24.53	100.00	<i>% of row</i>	75.47	24.53	100.00

Table 4.5: Possessive prefixes vs. other types (a) and possessive suffixes vs. other types (b) in the New World vs. the Old World.

Considering Eurasia separately presents a slightly different picture, although one that is somewhat unexpected. Table 4.6 on the next page presents the proportions of prefixes and suffixes in that area compared with the rest of the world.

	(a)			(b)			
	Eurasia	Other	Total	Eurasia	Other	Total	
Prefixes only	24	84	108	Suffixes only	11	53	64
<i>% of column</i>	<u>46.15</u>	<u>31.11</u>	33.54	<i>% of column</i>	21.15	19.63	19.88
<i>% of row</i>	22.22	77.78	100.00	<i>% of row</i>	17.19	82.81	100.00
Other	28	186	214	Other	41	217	258
<i>% of column</i>	53.85	68.89	66.46	<i>% of column</i>	75.85	80.37	80.12
<i>% of row</i>	13.08	86.92	100.00	<i>% of row</i>	15.89	84.11	100.00
Total	52	270	322	Total	52	270	322
<i>% of column</i>	100.00	100.00	100.00	<i>% of column</i>	100.00	100.00	100.00
<i>% of row</i>	16.15	83.85	100.00	<i>% of row</i>	16.15	83.85	100.00

Table 4.6: Possessive prefixes vs. other types (a) and possessive suffixes vs. other types (b) in Eurasia vs. other areas.

What we see is that in Eurasia the proportion of possessive prefixes is significantly higher than in the rest of the world ($p < 0.05$), sitting at nearly one-half instead of one-third. This observation runs contrary to Dryer’s claim that possessive prefixes are more prominent in the New World than in the Old World. But notice in Figure 4.5 where the majority of the datapoints are coming from. There is a cluster in the vicinity of the Himalayas. This area has been noted by Bickel and Nichols (2003) as a “typological enclave,” a small, usually secluded area that deviates from the larger area it belongs to. However, what is most interesting about this cluster is that most of the g-units are languages belonging to the Sino-Tibetan family. That family overall is significantly skewed toward having no possessive affixes, but 11 languages have possessive prefixes. These 11 languages belong to four genera of Sino-Tibetan: Bodic, Bodo-Garo, Karen, and Kuki-Chin. While it might be expected, since these genera are homogenous, that the g-sampling algorithm would count them at the genus level rather than the language level (4 g-units instead of 11), for some reason it has counted all of the languages. The reason may be that the genera are too small to pass the statistical test of significance.

Because of this artifact, it is illustrative to look at the counts again, this time reducing the deviating Sino-Tibetan genera to single points (effectively subtracting 7 from the “Prefixes only” category in Eurasia). The results are given in Table 4.7a on the following page.

	(a)			(b)			
	Eurasia	Other	Total	Eurasia	Other	Total	
Prefixes only	17	84	101	Suffixes only	11	53	64
<i>% of column</i>	<i>37.78</i>	<i>31.11</i>	<i>32.06</i>	<i>% of column</i>	<i>24.44</i>	<i>19.63</i>	<i>20.32</i>
<i>% of row</i>	<i>16.83</i>	<i>83.17</i>	<i>100.00</i>	<i>% of row</i>	<i>17.19</i>	<i>82.81</i>	<i>100.00</i>
Other	28	186	214	Other	34	217	251
<i>% of column</i>	<i>62.22</i>	<i>68.89</i>	<i>67.94</i>	<i>% of column</i>	<i>75.56</i>	<i>80.37</i>	<i>79.68</i>
<i>% of row</i>	<i>13.08</i>	<i>86.92</i>	<i>100.00</i>	<i>% of row</i>	<i>13.55</i>	<i>86.45</i>	<i>100.00</i>
Total	45	270	315	Total	45	270	315
<i>% of column</i>	<i>100.00</i>	<i>100.00</i>	<i>100.00</i>	<i>% of column</i>	<i>100.00</i>	<i>100.00</i>	<i>100.00</i>
<i>% of row</i>	<i>14.29</i>	<i>85.71</i>	<i>100.00</i>	<i>% of row</i>	<i>14.29</i>	<i>85.71</i>	<i>100.00</i>

Table 4.7: Possessive prefixes vs. other types (a) and possessive suffixes vs. other types (b) in Eurasia vs. other areas, with the genera of Sino-Tibetan reduced.

The distributions between areas are now much closer to each other and, unsurprisingly, not significant ($p \approx 0.39$). Even when we account for the artifact of genetic bias that we found, Eurasian languages do not appear to prefer possessive prefixes any less than languages elsewhere.

The distribution of suffixing g-units, shown in Table 4.6b, is once again not significantly differentiated by area ($p \approx 0.85$). Since we controlled for the Sino-Tibetan inflation effect when looking at prefixes, we do the same here for suffixes, subtracting 7 from the “Other” category in Eurasia. The results are given in Table 4.7b. Although the proportion of suffix-using g-units appears to be larger in Eurasia, the difference is still not significant ($p \approx 0.43$).

The overall result is that there is no significant areal difference in the distribution of possessive prefixes and possessive suffixes. That being established, there is a good possibility that possessive prefixes are preferred universally among languages that use pronominal possessive affixes (using the g-sampled numbers given in Table 4.4 on page 38, they make up 53.73 of the 201 g-units that use any sort of affixes). However, this possibility can be established for certain only when each macro-area is tested independently, as Dryer (1989) recommends.

It is also of interest to test this variable when all the positions of pronominal possessive affixes are grouped together, i.e. to compare languages with possessive affixes and those without. For this purpose, a reduced version of feature 57A was created, which we call 57A-r, in which the data was grouped together first and then run through the g-sampling algorithm. The division

based on Dryer’s original data is given in Table 4.8, and the breakdown of the g-sampled data is given in Table 4.8.

	Unsampled		G-sampled	
Any possessive affixes	642	71.17%	196	51.99%
No possessive affixes	260	28.82%	181	48.01%
Total	902	100.00%	377	100.00%

Table 4.8: Distribution of values for feature 57A-r: pronominal possessive affixes (reduced).

The difference between these two distributions is striking, as well as highly significant ($p < 0.005$). Whereas before the sampling possessive affixes had a large majority, g-sampling reveals the two options to be almost even. What is more, both values of this variable appear to be fairly evenly distributed throughout the world, as we see in Figures 4.7 and 4.8.

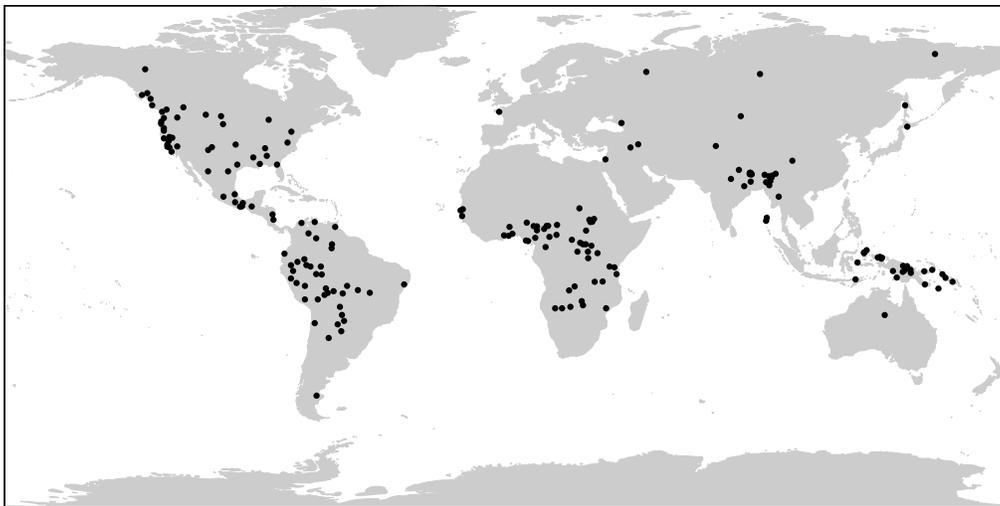


Figure 4.7: Geographical distribution of g-units with any pronominal possessive affixes.

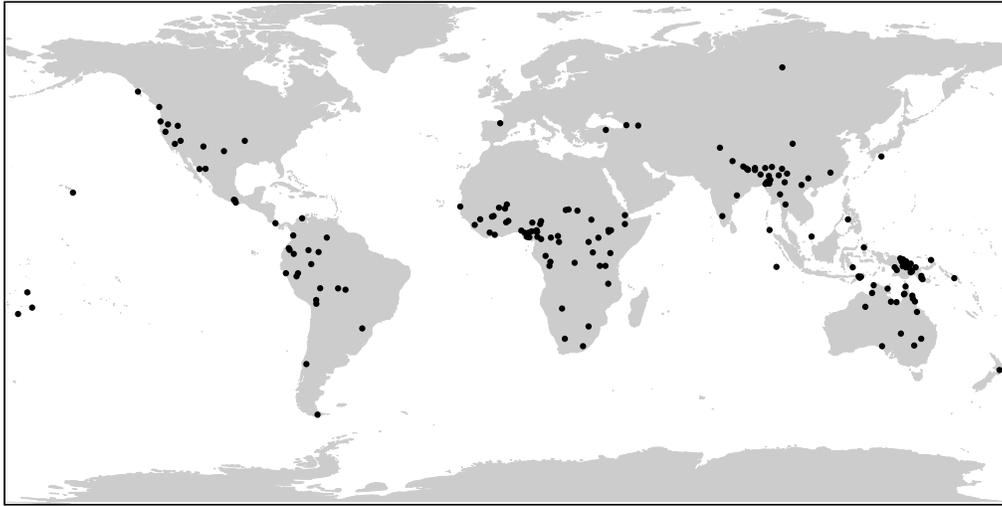


Figure 4.8: Geographical distribution of g-units with no pronominal possessive affixes.

Based solely on visual inspection, it does not seem likely that Eurasian languages are any different from the rest of the world with respect to possessive affixes. The numbers agree with this observation, as we see in Table 4.9.

	Eurasia	Elsewhere	Total
Any possessive affixes	33	163	196
<i>% of column</i>	<u>45.83</u>	<u>53.44</u>	51.99
<i>% of row</i>	16.84	83.16	100.00
No possessive affixes	39	142	181
<i>% of column</i>	54.17	46.56	48.01
<i>% of row</i>	20.42	78.45	100.00
Total	72	305	377
<i>% of column</i>	100.00	100.00	100.00
<i>% of row</i>	19.10	80.90	100.00

Table 4.9: Presence vs. absence of pronominal possessive affixes in Eurasia vs. other areas.

The distribution here is not significantly different from random ($p \approx 0.29$), meaning that the presence of possessive affixes is more or less equally as common as the lack of them in the

languages of Eurasia, just as it is in the rest of the world. Again, it cannot be said for certain whether this even split is the same everywhere in the world until each area is tested as Eurasia has been tested here, but we know at least that there is no strong preference in Eurasia with regard to this feature—nothing that makes it stand out.

Although we have found no geographic skewing overall, some families in the sampled have considerable skewing. Austronesian languages make up the largest group with skewing (112 languages), preferring possessive suffixes. A considerable number of Austronesian languages have no possessive affixes. Afro-Asiatic (57 languages) is also strongly skewed toward possessive suffixes, with a few genera in eastern Africa instead having no possessive affixes.

4.3 Obligatory possessive inflection

For feature 58A, there are only two possible values: either the language has obligatorily possessed nouns, or it does not. The left side of Table 4.10 shows the unsampled data from Bickel and Nichols’s (2011) survey. Figure 4.9 shows the geographical distribution.

	Unsampled		G-sampled	
Exists	43	17.62%	43	28.86%
Absent	201	83.38%	106	71.14%
Total	244	100.00%	149	100.00%

Table 4.10: Distribution of values for feature 58A: obligatory possessive inflection.

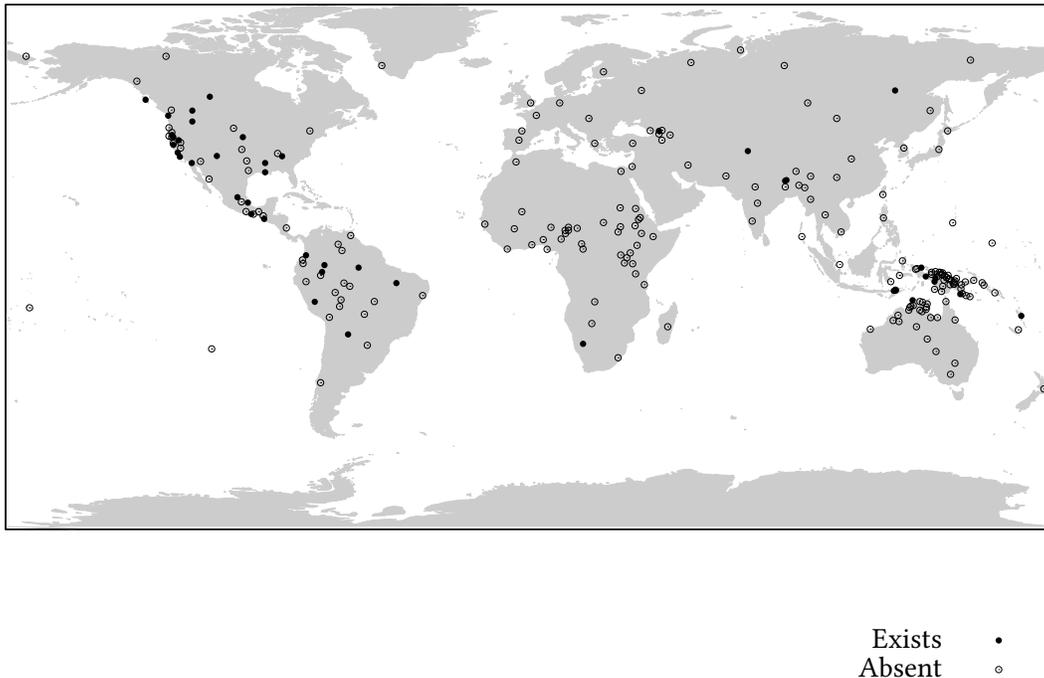


Figure 4.9: Geographical distribution of the unsampled data for obligatory possessive inflection.

It appears that obligatory possessive inflection is fairly rare across the world. However, the g-sampled data shows a different distribution, as we see in the right side of Table 4.10. The difference between these two sets of counts is significant ($p < 0.05$). It is notable that only languages without obligatorily possessed nouns are reduced through the g-sampling algorithm. They are still the more prominent type, but the number of g-units is just over half of the number of languages in the unsampled data. This difference suggests that obligatory possessive inflection is not as rare as it first appears, and is in fact fairly widespread among unrelated languages.

The geographical distribution of this feature is especially interesting. Figure 4.10 on the next page shows the distribution of g-units that have obligatory possessive inflection, and Figure 4.11 shows g-units that lack it. The difference in distribution between the two hemispheres is striking. The majority of g-units with obligatory possession are in the Americas, with only 6

appearing in Eurasia and Africa combined.² The visual data appears to corroborate Bickel and Nichols's (2011) argument that "Obligatorily possessed nouns are found chiefly in the Americas, where they are very common." G-units where obligatorily possessed nouns are absent, in contrast, appear to be more evenly distributed.

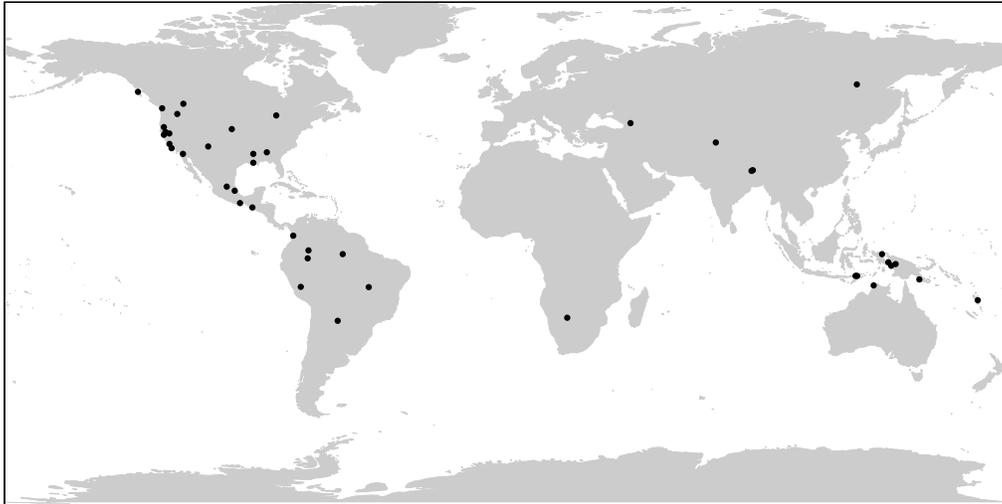


Figure 4.10: Geographical distribution of g-units with obligatory possessive inflection.

²Although it is not apparent on the map, Eurasia contains 5 g-units for this value. Belhare and Limbu, both Bodic languages spoken in Nepal, are so close together that they appear almost as a single point on this small map.

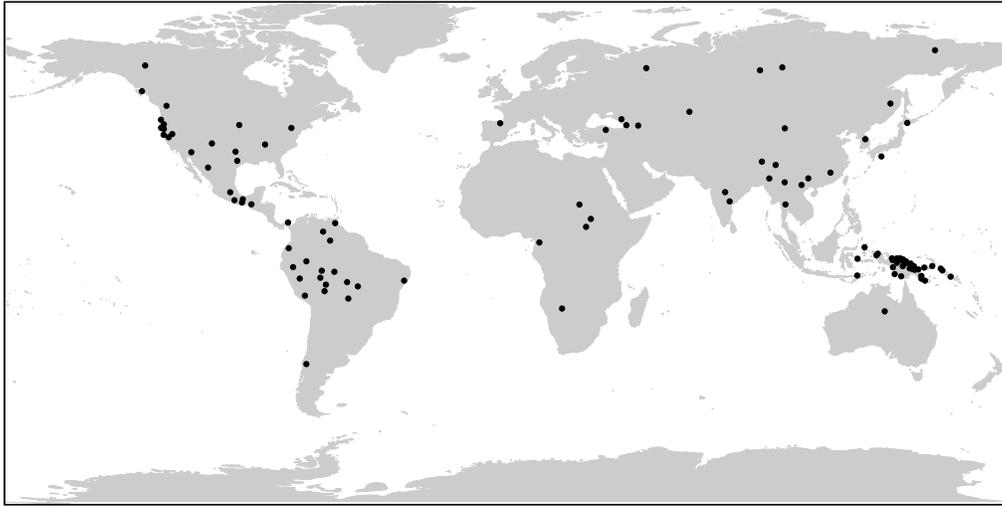


Figure 4.11: Geographical distribution of g-units without obligatory possessive inflection.

The numerical data supports this claim. Table 4.11 shows that obligatorily possessed nouns are twice as common in the Americas as they are everywhere else ($p < 0.01$). In addition, the row percentages indicate that nearly two-thirds of all g-units with obligatorily possessed nouns are found in the Americas. Obligatory possession appears to be a largely American feature.

	Americas	Elsewhere	Total
Exists	28	15	43
<i>% of column</i>	<u>40.00</u>	<u>18.99</u>	28.86
<i>% of row</i>	<u>65.12</u>	<u>34.86</u>	100.00
Absent	42	64	106
<i>% of column</i>	60.00	81.01	71.14
<i>% of row</i>	39.62	60.36	100.00
Total	70	79	149
<i>% of column</i>	100.00	100.00	100.00
<i>% of row</i>	46.96	53.02	100.00

Table 4.11: Presence vs. absence of obligatorily possessed nouns in the Americas vs. other areas.

It is less clear whether Eurasia significantly differs from the rest of the world in this feature. Although the data shown in Table 4.12 suggests that obligatory possession is much less common in Eurasia than in the rest of the world, the distribution fails the significance test ($p \approx 0.12$). This lack of significance probably stems from the fact that there are so few g-units in Eurasia for this feature. The sample contains several homogenous families in Eurasia, including Indo-European, Uralic, Dravidian, and Nakh-Daghestanian, all of which lack obligatory possessive inflection.

	Eurasia	Elsewhere	Total
Exists	5	38	43
<i>% of column</i>	<i>16.67</i>	<i>31.93</i>	<i>28.86</i>
<i>% of row</i>	<i>11.63</i>	<i>88.37</i>	<i>100.00</i>
Absent	25	81	106
<i>% of column</i>	<i>83.33</i>	<i>68.07</i>	<i>71.14</i>
<i>% of row</i>	<i>23.58</i>	<i>76.42</i>	<i>100.00</i>
Total	30	119	149
<i>% of column</i>	<i>100.00</i>	<i>100.00</i>	<i>100.00</i>
<i>% of row</i>	<i>20.13</i>	<i>79.87</i>	<i>100.00</i>

Table 4.12: Presence vs. absence of obligatorily possessed nouns in Eurasia vs. elsewhere.

The largest skewed families do not come from Eurasia, however. Australian is the largest family (25 languages) with significant skewing toward lacking obligatory possession, having only one deviating language, Tiwi. It is followed by Austronesian (15 languages), Niger-Congo (14 languages), and Nilo-Saharan (13 languages), all of which are also skewed toward lacking obligatory possession, the latter two uniformly. The vast majority of g-units in the Americas are singletons.

4.4 Possessive classification

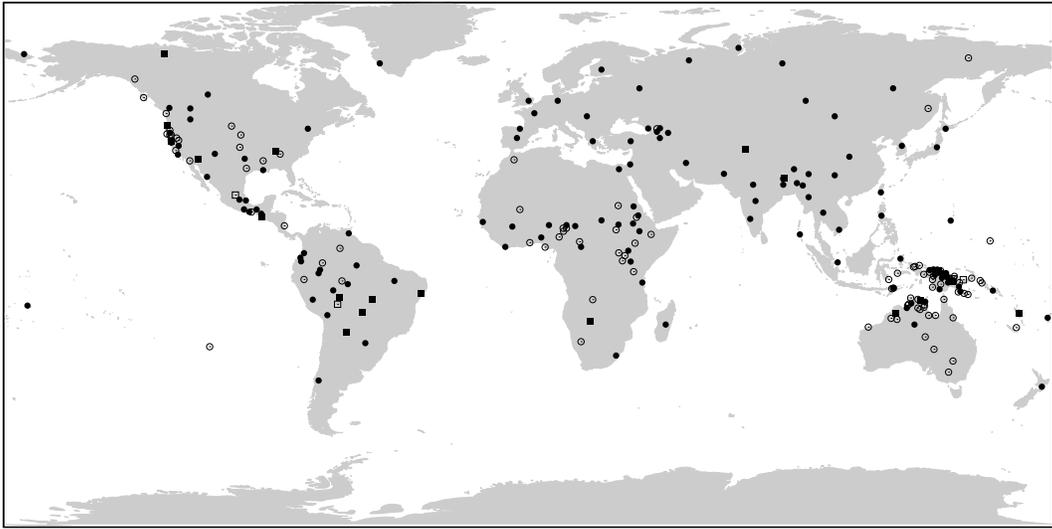
Nichols and Bickel (2011d) divide the data for feature 59A into four categories; they are shown along with their distributions in Table 4.13. Languages without any possessive classification make up the majority of the unsampled data.

	Unsampled		G-sampled	
No possessive classification	125	51.44%	77	44.77%
Two classes	94	38.68%	71	41.28%
Three to five classes	20	8.23%	20	11.63%
More than five classes	4	1.65%	4	2.33%
Total	243	100.00%	172	100.00%

Table 4.13: Distribution of values for feature 59A: possessive classification.

In the g-sampled data, however, the proportion of that category is considerably smaller. Both the zero category and the binary classification category are reduced to the point where they are almost equal. However, the difference between the sampled and unsampled distributions is neither significant ($p \approx 0.45$) when the table is taken as a whole, nor is it when only the two largest categories are considered ($p \approx 0.39$). Therefore, we can assume that there is no significant genetic bias in the data, and so either set can be used. For consistency, we will use the g-sampled data.

The zero category and the binary category are more or less equally common, at 44.77 and 41.28 percent respectively, and binary classification is the most common type of possessive classification, making up 74.74 percent of the three groupings of classification types. Having two classes is most common among languages with classification probably because of the notable alienability distinction discussed in Chapter 2. However, the maps reveal that the distribution is not even at all. Figure 4.12 on the following page shows the unsampled data, and Figures 4.13, 4.14, 4.15, and 4.16 show the g-sampled data for each individual value.



- No possessive classification •
- Two classes ◯
- Three to five classes ■
- More than five classes □

Figure 4.12: Geographical distribution of the unsampled data for possessive classification.

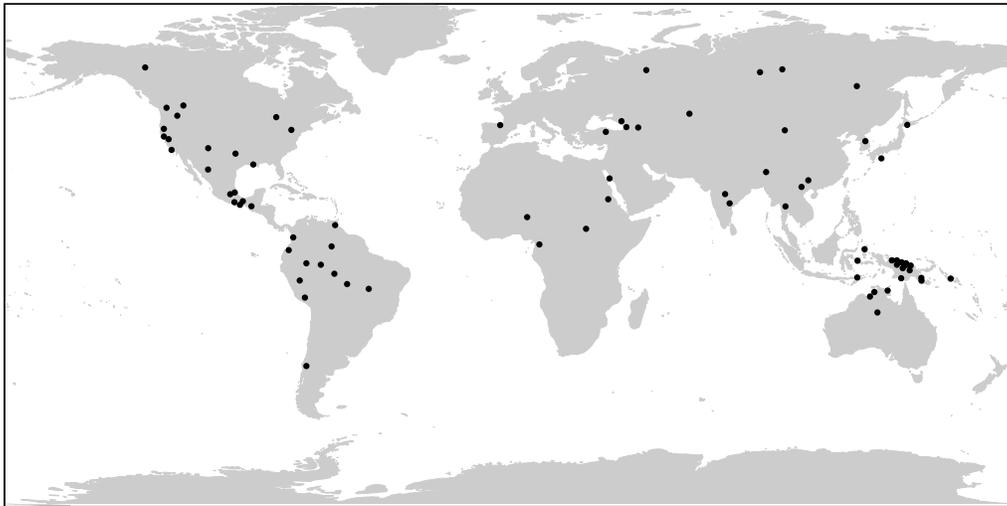


Figure 4.13: Geographical distribution of g-units with no possessive classification.

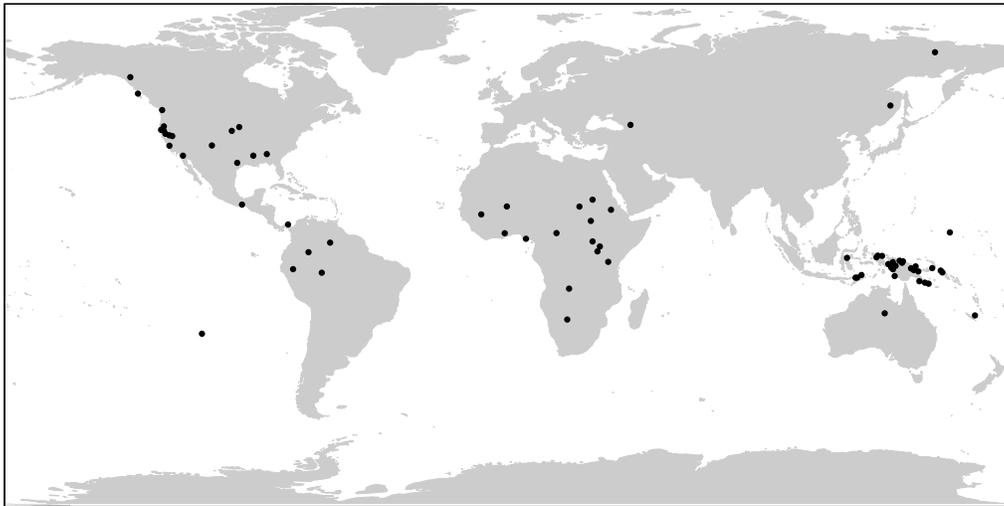


Figure 4.14: Geographical distribution of g-units with two possessive classes.

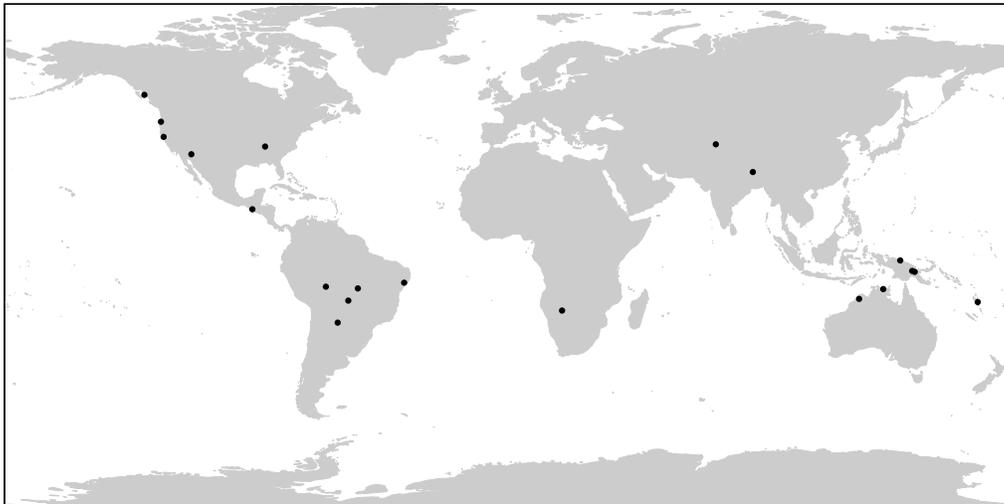


Figure 4.15: Geographical distribution of g-units with three to five possessive classes.

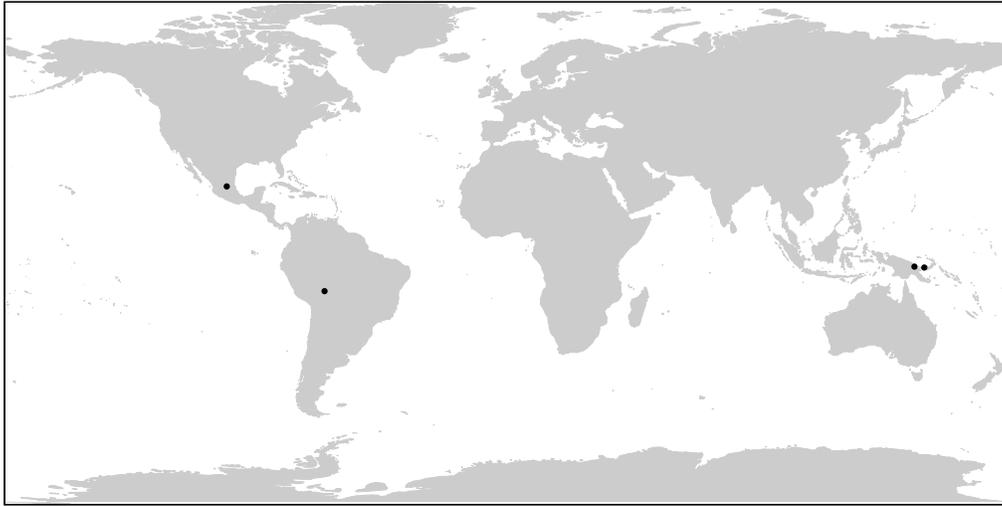


Figure 4.16: Geographical distribution of g-units with more than five possessive classes.

The Eurasian area differs strongly from this overall pattern, as shown in Table 4.14 ($p < 0.001$). Whereas g-units with two possessive classes are fairly common in the world in general (41.28 percent), in Eurasia they are somewhat rare (11.54 percent). And from the maps it is clear that binary classification is rare in Eurasia not because more complex systems are more common, but because possessive classification in general seems to be uncommon.

	Eurasia	Elsewhere	Total
Two classes	3	68	71
<i>% of column</i>	<i>11.54</i>	<i>46.58</i>	<i>41.28</i>
<i>% of row</i>	<i>7.23</i>	<i>95.77</i>	<i>100.00</i>
Other	23	78	101
<i>% of column</i>	<i>88.46</i>	<i>53.52</i>	<i>58.72</i>
<i>% of row</i>	<i>22.77</i>	<i>77.23</i>	<i>100.00</i>
Total	26	146	172
<i>% of column</i>	<i>100.00</i>	<i>100.00</i>	<i>100.00</i>
<i>% of row</i>	<i>15.12</i>	<i>84.88</i>	<i>100.00</i>

Table 4.14: Two possessive classes vs. other types in Eurasia vs. other areas.

To verify this claim, an alternative form of feature 59A was created, which we will call 59A-r. For this feature, all the languages that have any number of possessive classes were grouped into a single value, so that there were only two values, “exists” and “absent.” This reduced data was then run through the g-sampling algorithm. The distribution of the grouped original data, and the data after g-sampling, is shown in Table 4.15.

	Unsampled		G-sampled	
Exists	118	48.56%	89	46.60%
Absent	125	51.44%	102	53.40%
Total	243	100.00%	191	100.00%

Table 4.15: Distribution of values for feature 59A-r: possessive classification (reduced).

The difference between the sampled and unsampled data turns out not to be significant ($p \approx 0.7$), but as before we will use the g-sampled data. In general, languages are split almost evenly between having possessive classification and not having it. Once again, though, the geographical distribution is far from even, as we see in Figures 4.17 and 4.18. In every area except for Eurasia, g-units appear to be split fairly evenly between the two values, but in Eurasia, g-units without possessive classification vastly outnumber g-units with them.

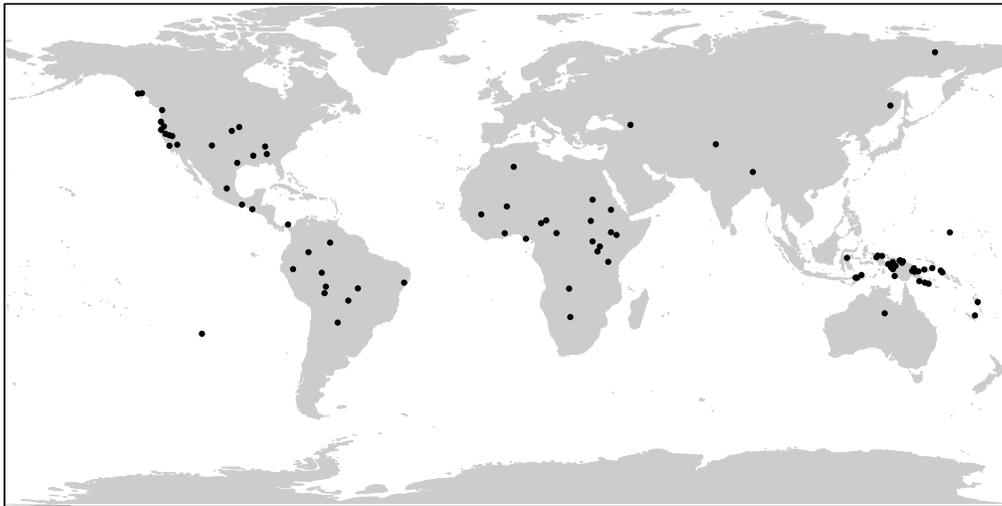


Figure 4.17: Geographical distribution of g-units that have possessive classification.

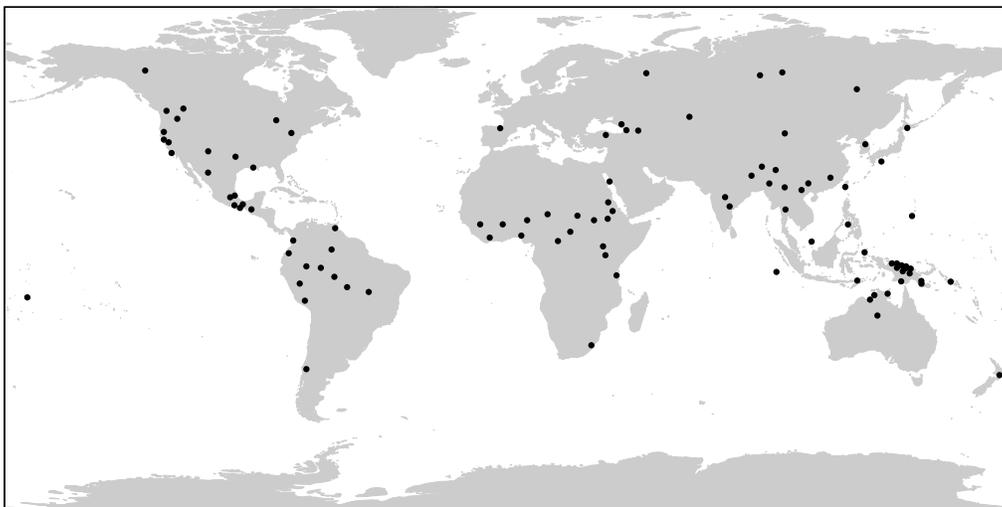


Figure 4.18: Geographical distribution of g-units with no possessive classes.

Visual inspection is confirmed by the actual counts shown in Table 4.16, which are highly significant ($p < 0.001$). The column percentages show that a mere 16.67 percent of g-units in

Eurasia use possessive classification, whereas in the rest of the world such g-units make up a slight majority at 53.55 percent. The row percentages give a different angle in showing that only 6.74 percent of g-units with possessive classification are found in Eurasia. This skewing is due partly to the fact that less than one-fifth (18.85 percent) of all the reported g-units come from Eurasia, but it cannot be explained completely by the uneven representation, since the row percentages for having classification are significantly different from the total row percentages. Possessive classification is rare in Eurasia.

	Eurasia	Elsewhere	Total
Exists	6	83	89
<i>% of column</i>	<u>16.67</u>	<u>53.55</u>	46.60
<i>% of row</i>	<u>6.74</u>	93.26	100.00
Absent	30	72	102
<i>% of column</i>	83.33	46.45	53.40
<i>% of row</i>	29.41	70.59	100.00
Total	36	155	191
<i>% of column</i>	100.00	100.00	100.00
<i>% of row</i>	<u>18.85</u>	81.15	100.00

Table 4.16: Presence vs. absence of possessive classification in Eurasia vs. elsewhere.

The largest skewed family for both 59A and 59A-r is once again Australian (25 languages), with a trend toward two possessive classes. For 59A, the next largest skewed families are Austronesian (15 languages), Niger-Congo (14 languages) and Nilo-Saharan (13 languages), all of which have a trend toward not having possessive classes. The second-largest skewed family for 59A-r is Indo-European, with 10 languages skewed toward not having possessive classes, as would be expected. The only deviate is Ossetic, an Iranian language spoken in Georgia that has two possessive classes.

4.5 Summary

In this chapter we have looked at the geographical distributions for four features of possessive morphology. Based on the data, we have made the following observations:

- Head marking is uncommon in the Old World. In the New World it is as common as all the other types combined.
- Dependent marking is common in the Old World and fairly uncommon in the New World.
- Eurasia holds roughly the same distribution as the Old World as a whole, except that head marking is even less common.
- There is no significant difference between the Old World and the New World with respect to the position of pronominal possessive affixes. Worldwide, only prefixes are found roughly one-third of the time, and only suffixes are found roughly one-fifth of the time.
- There is also no significant difference between Eurasia and the rest of the world with respect to the position of pronominal possessive affixes.
- Obligatory possessive inflection is not as rare as the unsampled data would suggest; over one-fourth of g-samples show the feature.
- Obligatory possessive inflection is twice as common in the Americas as in the rest of the world.
- Obligatory possessive inflection appears to be much less common in Eurasia than in the rest of the world, but the difference is not statistically significant.
- Binary possessive classification is much less common in Eurasia than in the rest of the world.
- Possessive classification in general is much less common in Eurasia than in the rest of the world.

In general, then, it appears that there are indeed quite a few differences between the Old and New Worlds with respect to possessive morphology, and those differences hold largely because of Eurasia. We will investigate possible reasons why these differences exist in Eurasia in Chapter 5.

The g-sampling method has also revealed that many large language families for which we have a lot of data are quite homogenous. The algorithm frequently groups together the Australian, Austronesian, Indo-European, Niger-Congo, and Nilo-Saharan families. Without proper sampling, the languages from these families can bias the results, which is why the g-sampling method proves so useful for making generalizations about the geographical distributions of linguistic features.

Chapter 5

Conclusions

5.1 The Eurasian macro-area revisited

We have confirmed two variables to be significantly different in Eurasia, namely, locus of marking in possessive noun phrases and possessive classification. A third variable, obligatory possessive inflection, is possibly different in Eurasia, but the difference is not statistically significant. The position of pronominal possessive affixes is no different between Eurasia and the rest of the world. These results differ slightly from those presented by Bickel and Nichols (2005, p. 15). They found locus of marking and possessive classification to be significantly different in Eurasia, as we have. However, they also found the presence of pronominal possessive affixes (the original feature reduced to a binary present-or-absent feature, like our 57A-r) to be significantly different in Eurasia ($p = 0.014$), whereas we did not ($p \approx 0.29$). They have made no data available on obligatory possessive inflection.

Now that we have seen where Eurasian languages are different, let us examine how they are different. In Eurasia, head marking is relatively rare (17.07 percent of g-units) and dependent marking is relatively common (43.90 percent of g-units), in direct contrast with the rest of the world. This fact is probably not surprising to speakers of Indo-European languages, since that family is skewed toward dependent marking. There is very little diversity in the whole region of Europe, but the diversity increases going eastward, with some examples of head marking as well as double marking. In Southeast Asia, there are quite a few more independent g-units, but dependent-marking units outnumber the rest. The feature is fairly widespread across the continent—at least the southern half. There are other types farther north, but they are few.

Of course, statistical distributions are not answers in themselves. Greenberg (1993, p. 505) himself stresses that “typological distributions are *explananda*, not explanatory principles.” The sampling and statistical tests performed in this study serve mainly as a way of identifying where

interesting areal patterns may be occurring, patterns that need to be explained. As to the explanatory principles themselves, there are many possibilities for distributions such as the ones we have found here. We explore them only briefly here.

Because there is a distinct geographical constraint on this preference for dependent marking, it cannot be explained simply by means of a structural universal. There is apparently no universal preference with regard to the locus of marking—no independent structural reason for one type to be preferable over another. There may, however, be connections to other structural features, like the correlations suggested by Greenberg (1963) and Dryer (1992, 2011c). In fact, one thing Nichols (1992) shows is that the locus of marking is a major predictor of other syntactic features, such as case alignment (e.g., ergative-absolutive vs. stative-active). However, she considers locus of marking as a whole feature of the language rather than only focusing on the possessive noun phrase as we do here.

It is useful to recall the informal observation from Nichols and Bickel (2011a) mentioned earlier in Section 2.3 of Chapter 2 that “the most common pattern for head-marked noun phrases” is when “the possessed noun (the head) agrees in person and number with the possessor noun.” In other words, the possessive marking often carries additional information besides merely encoding the possessive relationship. We saw this pattern with the Hungarian paradigm in example (14) of Chapter 2 (p. 15), repeated in part here for convenience:

- | | | | | |
|------|----|---------------|----------------|-----------|
| (22) | a. | a(z én) | ház-am | Hungarian |
| | | the (1SG.NOM) | house-1SG.POSS | |
| | | ‘my | house’ | |
| | b. | a (te) | ház-ad | |
| | | the (2SG.NOM) | house-1SG.POSS | |
| | | ‘your | house’ | |
| | c. | a (mi) | ház-unk | |
| | | the (1PL.NOM) | house-1PL.POSS | |
| | | ‘our | house’ | |
| | d. | a (ti) | ház-atok | |
| | | the (2PL.NOM) | house-2PL.POSS | |
| | | ‘your(PL) | house’ | |

Persian is a notable exception to this common pattern. Although it has a few different ways of encoding possession, one of them uses what is called an **ezafe** marker, which is affixed to the head of the phrase but encodes nothing more than the possessive relationship:

- (23) a. dom-e gorbe Persian (Mahootian, 1997, ex. 72–73)
 tail-EZ cat
 ‘the cat’s tail’
- b. doxtær-e mæn
 daughter-EZ I
 ‘my daughter’

However, the pattern exemplified by Hungarian is apparently more common, in which the possessive affix carries more grammatical meaning. In Dryer’s (2011b) terms, it is a pronominal possessive affix, and since they can occur only in head-marking languages, we might reasonably expect languages that use such affixes also to be considerably less common in Eurasia. The numbers (Table 4.9 on page 45), curiously, have shown that not to be the case, but it may be because the databases for the two features are vastly different in size.

The pattern of possessive classification may possibly be correlated with the locus of marking. We have shown possessive classification to be rare in Eurasia and fairly common everywhere else. It would make sense for possessive classes to be less common in an area where head-marking is less common because possessive classification is a way of lexically categorizing the possessum—the head of the phrase. It is true that sometimes the ways of indicating the different possessive classes do not always involve directly marking the head, as we saw in the example from Ngitì in Chapter 2:

- (24) a. kamà bhà dza Ngitì
 chief GEN house
 ‘the chief’s house’
- b. kamà-dò
 chief-head
 ‘the chief’s head’

But even in this example, the inalienable noun takes direct marking. Most examples of possessive classification occur in head-marking contexts. What does this mean, then? There seems to be something going on here with respect to the level of specificity in the possessive

construction. On the one hand, we have languages that encode possession simply by juxtaposing nominals, thus encoding nothing specific about the relationship at all. Then we have languages that use markers that encode nothing more than a relationship, sometimes with a generic linking marker as in Chamorro (14), and sometimes with a genitive marker as English uses. But then on the other end of the spectrum are languages that encode quite a bit more detail in the possessive construction, showing morphological agreement and sometimes classification. And it seems, based on this brief survey, that head-marking languages are more likely to encode more specific information in the possessive noun phrase than are dependent-marking languages. Haspelmath (1993a, p. 496) argues that the reason for this association is a problem with the definition of the head-dependent marking distinction as put forth by Nichols (1986, 1992). Rather than head marking, Haspelmath says it should be called agreement marking, and dependent marking should be called case marking, because that seems to be the real distinction that is occurring. This issue merits further investigation.

If there are structural connections between features, such as between the locus of marking and possessive classification as we have speculated here, it is difficult to see if one came before the other and thus “caused” the other because they are structurally compatible (in other words, it is a match feature as discussed in Chapter 1). But there are other factors that can determine the preference of one type over another, and since we have shown Eurasia as a linguistic area to show unique preferences, it is important to consider factors related to the geography. This is why Bickel and Nichols (2005, p. 2) assert that it is important to ground areal patterns in theories of population history, to consider languages not only in terms of their abstract construction but also in terms of the populations that speak them.

It would be beyond the scope of this paper to delve into the histories of language contact and movement that has occurred in Eurasia, but it is perhaps instructive to touch briefly at least on the size of the language populations we are dealing with here. Some of the language families with the largest numbers of native speakers are found in Eurasia: Indo-European, Sino-Tibetan, Dravidian, and Altaic. As it so happens, many of these families are typologically similar. Specifically, Indo-European, Sino-Tibetan, and Dravidian are all skewed toward dependent marking and lacking possessive classification. Since there are so many speakers of languages of these types, it is certainly possible that these languages would have an influence on languages they come in con-

tact with, especially in cases where speakers of smaller native languages need to learn the larger, perhaps national language for economic or political reasons. Features that are spread from the majority to minorities are spread features as described in Chapter 1. It is conceivable that the features we have discussed here were spread from these large language families to smaller families simply because they have more native speakers.

All of these possible reasons for the typological differences in Eurasian languages are admittedly mere speculation. At the very least, though, the statistics found in this study provide a starting point for deeper investigations into the histories of feature spreads and the reasons behind them.

5.2 Possession as a general feature

One interesting question that our results here raise is whether individual features of possession tend to be diffused as bundles. We have looked at several specific features of possessive morphology and have found that at least two of them have unique behavior in Eurasia (the other two may also be different in Eurasia, but the statistics here do not confirm that). The fact that the behavior of these features is affected by roughly similar geographic boundaries suggests that not only might they have been influenced by the same historical, sociological, or anthropological factors, but also they may be related to each other in a way that makes it likely for them to be transferred together. We have already seen above, for example, how the locus of marking can affect other features. These interrelationships raise the question of whether possession can be considered a cohesive subdomain of morphosyntax, in which the individual features (locus of marking, classification, obligatory inflection, etc.) tend to go together and are diffused into different languages together, rather than just general features of morphosyntax that happen to relate to possession.

This line of thinking perhaps contradicts Bickel's (2007) demand for fine-grained variables. It also raises problems of cross-linguistic comparability, which Haspelmath (2007) argues is hampered by the use of predetermined categories. Indeed, general observations cannot be made until the specifics are worked out; Bickel and Nichols (2005, p. 2) argue that "each variable can reflect areal factors on its own terms." But if it does actually turn out that many fine-grained variables

dealing with possession display similar areal effects, it may be worth investigating whether the variables are in fact structurally or implicationally bundled.

It is true that, in the case of this study, only two variables out of the eight in WALS that deal with possession were confirmed to be different in Eurasian languages. While that proportion is not particularly impressive, it is important to remember that any single variable that shows areal effects is of interest, because there are areal causes to investigate. When even two variables coincide areally, as they largely have done in this study, there is even more to investigate: the connection between the variables, and whether that connection is structural or tied to the geography. It would certainly be a jump to assume that all possessive features have similar areal patterns, but the possibility cannot yet be ruled out, because of the behaviors discovered in this study.

5.3 Limitations and possibilities

Although the results presented in this paper can be considered reliable, there are some limitations to the methods used here that need to be acknowledged. First, in addition to the problems inherent to language sampling already discussed, which we attempted to minimize, our samples depended on a pre-established genealogy of languages. But genealogical classifications of languages can be controversial, especially when it comes to smaller, “exotic” languages we know little about. Cysouw (2005, p. 556), among others, warns that the classification system used to base a language sample on can dramatically affect the results of a typological study. In this study, because we used data directly from WALS, we had to use the genealogical groupings used in WALS. It is possible that the relationships in that database are not entirely accurate, or at least not detailed enough—most of the languages are organized into only two levels of groups. More detailed grouping levels are certainly conceivable, and they could result in quite different outputs from the g-sampling algorithm.

On a related note, we are limited of course by how much data is available in WALS. As discussed before in Chapter 3, not every language is documented with the same thoroughness or even in the same way. Furthermore, each feature uses a different sample of languages, based on what the contributing author could find in the available documentation. It would be ideal if

every feature had data on the same languages, thus allowing us to make real comparisons between features. A simply larger volume of documented languages would also be helpful.

One weakness of the method used in this study is the way that higher-level g-units are treated. As noted in Section 3.4 in Chapter 3, the coordinates for genera and families were calculated by averaging the coordinates of their member languages. Such averages can create a somewhat imprecise picture of the locations of g-units, particularly when a group covers a large geographical area. However, since this study is looking at a large macro-area of languages, these imprecise locations were deemed to be less of a problem than they would be if we were looking at smaller areas, or if the statistical tests depended directly on the coordinates themselves, which they did not.

These limitations aside, the goal of this project has been to show that the languages of Eurasia are significantly different from the rest of the languages of the world with respect to features of possessive morphology. Now that some significant differences have been shown, the next step is to find definitive reasons behind these areal patterns. Of particular interest is discovering if and how these features relate to each other, as well as tracing their history of areal influences—whether they came from majority languages in the area or whether there is some other reason they are so prevalent.

This study has also shown a method for identifying areal patterns that can be used for other features. The WALS database has a large number of features in a wide variety of topics, and it would be interesting to see if this method could find or confirm other areal patterns that these features follow. Of course, it is not limited to WALS data but could be used with any typological database with a sufficient amount of data and reasonably accurate and detailed language classification.

References

- Abbott, M. (1991). Macushi. In D. C. Derbyshire & G. K. Pullum (Eds.), *Handbook of Amazonian languages: Vol. 3*. (Vols. 3, pp. 23–160). Berlin: Mouton de Gruyter.
- Bakker, D. (2011). Language sampling. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology* (pp. 100–127). Oxford: Oxford University Press.
- Bell, A. (1978). Language samples. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language: Vol. 1*. (pp. 123–156). Stanford: Stanford University Press.
- Bickel, B. (2002). *The AUTOTYP research program*. Invited talk given at the Annual Meeting of the Linguistic Typology Resource Center, Utrecht, Netherlands, September 26–28, 2002. Retrieved from <http://www.spw.uzh.ch/autotyp/download/ltrc02.pdf>
- Bickel, B. (2007). Typology in the 21st century: major current developments. *Linguistic Typology*, 11, 239–251. doi:10.1515/LINGTY.2007.018
- Bickel, B. (2008). A refined sampling procedure for genealogical control. *Sprachtypologie und Universalienforschung*, 61(3), 221–233.
- Bickel, B. (2011). Statistical modeling of language universals. *Linguistic Typology*, 15, 401–413. doi:10.1515/LITY.2011.027
- Bickel, B. & Nichols, J. (2003). *Typological enclaves*. Presented at the 5th Biannual Conference of the Association for Linguistic Typology, Cagliari, Italy, September 18, 2003. Retrieved from <http://www.spw.uzh.ch/autotyp/download/enclaves@ALT5-2003BB-JN.pdf>
- Bickel, B. & Nichols, J. (2005). *Areal patterns in the World Atlas of Language Structures*. Presented at the 6th Biannual Conference of the Association for Linguistic Typology, Padang, Indonesia, July 24, 2005. Retrieved from http://www.spw.uzh.ch/autotyp/download/areas_wals@alt6final.ppt.pdf
- Bickel, B. & Nichols, J. (2007). Inflectional morphology. In T. Shopen (Ed.), *Language typology and syntactic description: Vol. 3*. (2nd ed., pp. 169–240). Cambridge: Cambridge University Press.

- Bickel, B. & Nichols, J. (2011). Obligatory possessive inflection. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chap. 58). Munich: Max Planck Digital Library. Retrieved from <http://wals.info/chapter/58>
- Campbell, L., Kaufman, T., & Smith-Stark, T. C. (1986). Meso-America as a linguistic area. *Language*, 62(3), 530–570.
- Chapman, S. & Derbyshire, D. C. (1991). Paumari. In D. C. Derbyshire & G. K. Pullum (Eds.), *Handbook of Amazonian languages: Vol. 3*. (Vols. 3, pp. 161–352). Berlin: Mouton de Gruyter.
- Comrie, B. & Smith, N. (1977). Lingua descriptive studies: Questionnaire. *Lingua*, 42(1), 1–72.
- Croft, W. (1990). *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge: Cambridge University Press.
- Cysouw, M. (2005). Quantitative methods in typology. In R. Köhler, G. Altmann, & R. G. Piotrowski (Eds.), *Quantitative linguistics: An international handbook* (pp. 554–578). Handbooks of Linguistics and Communication Science. Berlin: Mouton de Gruyter.
- Daniel, M. (2011). Linguistic typology and the study of language. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology* (pp. 42–68). Oxford: Oxford University Press.
- Dryer, M. S. (1989). Large linguistic areas and language sampling. *Studies in Language*, 13(2), 257–292.
- Dryer, M. S. (1992). The Greenbergian word order correlations. *Language*, 68(1), 81–138.
- Dryer, M. S. (1997). Why statistical universals are better than absolute universals. In K. Singer (Ed.), *Papers from the regional meetings* (Vol. 33, pp. 123–145). Chicago Linguistic Society.
- Dryer, M. S. (2000). Counting genera vs. counting languages. *Linguistic Typology*, 4(3), 334–356.
- Dryer, M. S. (2003). Significant and non-significant implicational universals. *Linguistic Typology*, (1), 108–128.
- Dryer, M. S. (2007). Noun phrase structure. In T. Shopen (Ed.), *Language typology and syntactic description: Vol. 2*. (2nd ed., pp. 151–205). Cambridge: Cambridge University Press.
- Dryer, M. S. (2009). Problems testing typological correlations with the online WALS. *Linguistic Typology*, 13, 121–135. doi:10.1515/LITY.2009.007
- Dryer, M. S. (2011a). Order of genitive and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chap. 86). Munich: Max Planck Digital Library. Retrieved from <http://wals.info/chapter/86>

- Dryer, M. S. (2011b). Position of pronominal possessive affixes. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chap. 57). Munich: Max Planck Digital Library. Retrieved from <http://wals.info/chapter/57>
- Dryer, M. S. (2011c). The evidence for word order correlations. *Linguistic Typology*, 15, 335–380. doi:10.1515/LITY.2011.024
- Dryer, M. S. & Haspelmath, M. (Eds.). (2011). *The world atlas of language structures online*. Munich: Max Planck Digital Library, retrieved from <http://wals.info>
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of human language* (pp. 73–113). Cambridge, Mass: MIT Press.
- Greenberg, J. H. (1993). [Review of *Linguistic diversity in space and time* by J. Nichols, 1992]. *Current Anthropology*, 34(4), 503–505
- Haspelmath, M. (1993a). [Review of *Linguistic diversity in space and time* by J. Nichols, 1992]. *Journal of Linguistics*, 29(2), 494–500
- Haspelmath, M. (1993b). *A grammar of Lezgian*. Berlin: Mouton de Gruyter.
- Haspelmath, M. (2007). Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology*, 11(1), 119–132. doi:10.1515/LINGTY.2007.011
- Herslund, M. & Baron, I. (2001). Introduction: Dimensions of possession. In I. Baron, M. Herslund, & F. Sørensen (Eds.), *Dimensions of possession* (Vol. 47, pp. 1–26). Typological Studies in Language. Amsterdam: John Benjamins.
- Janssen, D. P., Bickel, B., & Zúñiga, F. (2006). Randomization tests in language typology. *Linguistic Typology*, 10(3), 419–440. doi:10.1515/LINGTY.2006.013
- Koptjevskaja-Tamm, M. (2002). Adnominal possession in the European languages: Form and function. *Sprachtypologie und Universalienforschung*, 55(2), 141–172.
- Koptjevskaja-Tamm, M. (2003). Possessive noun phrases in the languages of Europe. In F. Plank (Ed.), *Noun phrase structure in the languages of Europe* (Vol. 20-7, pp. 621–722). Empirical Approaches to Language Typology. Berlin: Mouton de Gruyter.
- Kuipers, A. H. (1967). *The Squamish language, grammar, texts, dictionary*. Janua Linguarum, Series Practica. The Hague: Mouton.

- Kutsch Lojenga, C. (1994). *Ngiti, a Central-Sudanic language of Zaire*. Cologne: Rüdiger Köppe Verlag.
- Lastra de Suárez, Y. (1984) In M. S. Edmonson (Ed.), *Supplement to the handbook of Middle American languages: Linguistics* (pp. 20–42). Austin: University of Texas Press.
- Lewis, M. P. (Ed.). (2009). *Ethnologue: Languages of the world*. Dallas: SIL International, retrieved from <http://www.ethnologue.com>
- Mahootian, S. (1997). *Persian*. London: Routledge.
- Manzelli, G. (1990). Possessive adnominal modifiers. In J. Bechert, G. Bernini, & C. Buridant (Eds.), *Toward a typology of European languages* (pp. 63–111). Berlin: Mouton de Gruyter.
- Miller, W. R. (1965). *Acoma grammar and texts*. University of California Publications in Linguistics. Berkeley: University of California Press.
- Nedjalkov, I. (1997). *Evenki*. London: Routledge.
- Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, 62(1), 56–119.
- Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Nichols, J. & Bickel, B. (2011a). Locus of marking in possessive noun phrases. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chap. 24). Munich: Max Planck Digital Library. Retrieved from <http://wals.info/chapter/24>
- Nichols, J. & Bickel, B. (2011b). Locus of marking in the clause. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chap. 23). Munich: Max Planck Digital Library. Retrieved from <http://wals.info/chapter/23>
- Nichols, J. & Bickel, B. (2011c). Locus of marking: whole-language typology. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chap. 25). Munich: Max Planck Digital Library. Retrieved from <http://wals.info/chapter/25>
- Nichols, J. & Bickel, B. (2011d). Possessive classification. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chap. 59). Munich: Max Planck Digital Library. Retrieved from <http://wals.info/chapter/59>
- Payne, D. L. & Barshi, I. (Eds.). (1999). *External possession*. Amsterdam: John Benjamins, Typological Studies in Language.
- Perkins, R. (1989). Statistical techniques for determining language sample size. *Studies in Language*, 13(2), 293–315.

- Perkins, R. (2001). Sampling procedures and statistical methods. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: An international handbook* (Chap. 33, Vol. 1, Vols. 2, pp. 419–434).
- Plank, F. (1995). (Re-)introducing Suffixaufnahme. In F. Plank (Ed.), *Double case: Agreement by Suffixaufnahme* (Chap. 1, pp. 3–110). New York: Oxford University Press.
- Rijkhoff, J. (2002). *The noun phrase*. Oxford: Oxford University Press.
- Rijkhoff, J., Bakker, D., Hengevald, K., & Kahrel, P. (1993). A method of language sampling. *Studies in Language*, 17(1), 169.
- Seiler, H. (1983). *Possession as an operational dimension of language*. Language Universals Series. Tübingen: Gunter Narr.
- Sridhar, S. N. (1990). *Kannada*. London: Routledge.
- Stassen, L. (2009). *Predicative possession*. Oxford: Oxford University Press.
- Stassen, L. (2011). Predicative possession. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chap. 117). Munich: Max Planck Digital Library. Retrieved from <http://wals.info/chapter/117>
- Topping, D. M. (1980). *Chamorro reference grammar*. Honolulu: University of Hawaii Press. (Original work published 1973)
- Tucker, A. N. & Bryan, M. A. (1966). *Linguistic analysis: the non-Bantu languages of north-eastern Africa*. London: Oxford University Press.
- Ultan, R. (1978). Toward a typology of substantival possession. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language: Vol. 4*. (pp. 11–49). Stanford: Stanford University Press.
- van Driem, G. (1993). *A grammar of Dumi*. Berlin: Mouton de Gruyter.
- Young, R. W. & Morgan, W. (1987). *The Navajo language: A grammar and colloquial dictionary* (Rev. ed.). Albuquerque: University of New Mexico Press.