Historical Linguistics

An Introduction

SECOND EDITION

Lyle Campbell

The MIT Press
Cambridge, Massachusetts

HP020364

these to nasals, nothing would follow for subgrouping, since Nootka alone would change and Makah and Nitinat would only share a retention.

The Mayan subgrouping, considered above, provides a final example, though it is simplified here in that we will consider only one of many sound correspondences together with the changes and the reconstruction based on it. In the Mayan family, the lower-level subgroups are well established; these include Huastecan, Yucatecan, Cholan-Tzeltalan, Greater Q'anjobalan, K'ichean and Mamean. Some of these are grouped together in higher-order, more inclusive branches of the family; we must ask what the evidence for these larger subgroupings is and whether it is accurate. Consider the following sound correspondence (encountered earlier, in Chapter 5):

Huastecan *h* : Yucatecan *n* : Cholan-Tzeltalan *n* : Q'anjobalan *ŋ* : K'ichean *x* : Mamean *x*

The generally accepted reconstruction in this case is Proto-Mayan **ŋ* (where it is assumed that Huastecan independently changed **ŋ > h*, change (3) in the list above), and so we will leave it out of the rest of the discussion). K'ichean and Mamean share the change of **ŋ > x* (change (12) above; *x* then later changed to *x* in Mamean and in most of the K'ichean languages), and this shared innovation (together with others mentioned above) supports subgrouping K'ichean and Mamean together; the group is usually called Eastern Mayan. In this reconstruction for the correspondence set that Proto-Mayan **ŋ* is based on, Yucatecan, Cholan-Tzeltalan and Q'anjobalan each retain the nasal (where it is assumed that the change of *ŋ > n* is so natural and easy that Yucatecan and Cholan-Tzeltalan probably underwent it independently), and since this is a shared retention (if viewed this way), nothing follows for whether or not these three groups may have a closer kinship or not. However, K'ichean and Mamean share the innovation **ŋ > x*, which is grounds for subgrouping them together. Suppose hypothetically now that this reconstruction were wrong and that Proto-Mayan actually had **x* (although this is highly unlikely). In this case, K'ichean and Mamean would share not an innovation but merely a retention, and nothing would follow from this for their position within the family. However, Yucatecan, Cholan-Tzeltalan and Greater Q'anjobalan would all share an innovation to a nasal (**x > ŋ*, then later *ŋ > n* in Yucatecan and Cholan-Tzeltalan), and this would be evidence for classifying Yucatecan, Cholan-Tzeltalan and Greater Q'anjobalan as members of the same subgroup. That is, if the reconstruction of **ŋ* is wrong, then the subgrouping based on the shared innovations which depart from this

reconstruction is also not founded; if the alternative reconstruction with **x* is wrong (which is almost certainly the case), then any subgrouping which presupposes it must also be wrong (unless other shared innovations can be found which do support it).

6.5 Glottochronology (Lexicostatistics)

Not all methods of classification that have been proposed are reliable. Glottochronology is a well-known one which is still sometimes used but which has been rejected by most historical linguists. In what follows, it is discussed in some detail, not because it merits such attention, but because it has proven particularly misleading and it is important to understand why it should be avoided. It is sometimes likened to ^{14}C ('carbon 14') dating in archaeology. Given the attention it has received (and continues to receive in some quarters), it is important to understand why it does not work for subgrouping, or for any other purpose, for that matter.

Though the names *glottochronology* and *lexicostatistics* are usually used interchangeably, some make a distinction; *glottochronology* is defined as a method with the goal of assigning a date to the split-up of some language into daughter languages, whereas *lexicostatistics* is given the definition of the statistical manipulation of lexical material for historical inferences (not necessarily associated with dates). Lexicostatistics in this sense is broader. However, in actual practice, this distinction is almost never made; both names are used interchangeably.

6.5.1 Basic assumptions

There are four basic assumptions of glottochronology, all of which have been challenged. We will look at each in turn and consider some of the criticisms that have been raised concerning them.

(1) *Basic vocabulary*. The first assumption is that there exists a *basic* or *core vocabulary* which is universal and relatively culture-free, and thus is less subject to replacement than other kinds of vocabulary. The Swadesh 100-word list of basic vocabulary is:

1. I	7. who	13. big	19. fish
2. you	8. not	14. long	20. bird
3. we	9. all	15. small	21. dog
4. this	10. many	16. woman	22. louse
5. that	11. one	17. man	23. tree
6. what	12. two	18. person	24. seed

HP020365

25. leaf	44. tongue	63. swim	82. fire
26. root	45. claw	64. fly	83. ash
27. bark	46. foot	65. walk	84. burn
28. skin	47. knee	66. come	85. path
29. flesh	48. hand	67. lie	86. mountain
30. blood	49. belly	68. sit	87. red
31. bone	50. neck	69. stand	88. green
32. egg	51. breast	70. give	89. yellow
33. grease	52. heart	71. say	90. white
34. horn	53. liver	72. sun	91. black
35. tail	54. drink	73. moon	92. night
36. feather	55. eat	74. star	93. hot
37. hair	56. bite	75. water	94. cold
38. head	57. see	76. rain	95. full
39. ear	58. hear	77. stone	96. good
40. eye	59. know	78. sand	97. new
41. nose	60. sleep	79. earth	98. round
42. mouth	61. die	80. cloud	99. dry
43. tooth	62. kill	81. smoke	100. name

To apply glottochronology, lists of the most natural, most neutral translations of each of these 100 semantic concepts are assembled and compared in two or more related languages – or at least languages thought to be related. The forms which are phonetically similar in the compared lists receive a check mark (tick) to indicate probable cognates; and, as will be seen below, the date when these languages separated from one another is calculated based on the number of these checked/ticked 'cognates' that they share. Some scholars argue that the method should be constrained to require that only forms known from historical linguistic research to be real cognates be counted, rather than mere 'look-alikes', as in the more common practice.

(2) *Constant rate of retention through time*. The second assumption is that the *rate of retention* of items of core vocabulary is relatively constant through time, that a language will retain about 86 per cent of the words of the 100-word list each 1,000 years (the figure is 80.5 per cent, rounded to 81 per cent retention for the 200-word list, formerly used but found not to be sufficiently culture-free and therefore replaced by the 100-word list).

(3) *Constant rate of loss cross-linguistically*. The third assumption is related to the second; it claims that the *rate of loss* of basic vocabulary is approximately the same for all languages. It is assumed that languages everywhere lose about 14 per cent of the 100-word list, that is, that some

fourteen words from the 100-word list will be replaced (thus some eighty-six of the basic 100 words will be retained) each 1,000-year period throughout their history.

(4) *Calculation of the date of divergence*. The fourth assumption is that when the number of cognates in the basic vocabulary list shared by related languages is known, the number of centuries since the languages split from an earlier ancestor can be computed. The time depth is computed with the formula

$$t = \frac{\log C}{2 \log r}$$

where t is 'time depth' in millennia (1,000-year periods), C is 'percentage of cognates' and r is 'the constant' (the percentage of cognates assumed to remain after 1,000 years, that is, 86 per cent for the 100-word list). *Log* means 'logarithm of'.

6.5.2 Historical background

Glottochronology was invented in the 1950s by Morris Swadesh, an American linguist, who began by trying to determine whether there were broad trends involving vocabulary change within particular language families. He was surprised to discover, so he reported, that not only were there constant trends within particular language families, but that the rate of change turned out to be the same across languages, regardless of their family affiliations. This claim constitutes one of the basic assumptions of the method, and it has been vigorously criticized (see below). Swadesh began with a basic vocabulary list of 500 words, but this was soon reduced to 205, then to 200, and finally to the 100-word list. In developing glottochronology, he examined thirteen test cases – languages with long attested histories where vocabulary change could be checked against written evidence. In these 'test cases', he compared modern versions of English, German and Swedish (Germanic languages) with older attested stages of each language (for example, Modern English with Old English). Catalan, French, Italian, Portuguese, Romanian and Spanish (Romance languages) were compared with Latin. Athenian Greek and Cypriotic Greek were compared with Classical Greek; Coptic was compared with Middle Egyptian (its ancestor); and modern Mandarin Chinese was compared with Ancient Chinese. However, only two of these thirteen (Coptic and Mandarin) are non-Indo-European languages, and this has raised doubts about the method. From later tests with control cases involving Kannada, Japanese, Arabic, Georgian, Armenian

HP020366

and Sardinian, the claim of a constant rate of retention has been challenged (see below).

6.5.3 Criticisms

6.5.3.1 Problems with the assumption of basic vocabulary

There are serious problems with the assumption of a universal, culture-free basic vocabulary. One is that many of the items are not culture-free, but rather are borrowed for cultural reasons in a number of languages. Examples of borrowed terms for items on the 100-word list are found for each item in some language somewhere; only a few revealing examples are mentioned here. In several Mayan languages, (18) *winaq* 'person' was replaced by a loanword, *kriftian* (or something similar), from Spanish *cristiano* 'Christian', colloquially 'person, living soul'. It is thought that in the early colonial period, Spanish contrasted Christianized Indians (the *cristianos*) with pagans. When ultimately all had been 'pacifized' (converted), by default all were then called *kriftian* 'person', resulting in the elimination from the vocabulary of former *winaq* 'person'. In the case of (21) 'dog', while native peoples of Central America had dogs before the coming of the Spanish, their dog was small, hairless and barkless, and served as a food item. The big, hairy, noisy dogs which arrived with Europeans were not easily equated with the native dogs, and hence many groups borrowed the foreign name for 'dog' and eventually came no longer to have a native term for 'dog'. Thus, for example, 'dog' in Pipil (Uto-Aztecan) is *pe:lu*, borrowed from Spanish *perro* 'dog' (Pipil has no *r*). The word for (52) 'heart' is widely borrowed in a number of Mayan languages from Totonac (a non-Mayan language of Mexico); this presumably has to do with the importance of 'heart' in native religion (for example, human sacrifice by cutting out the heart was practised). Forms for (72) 'sun' and (73) 'moon' are widely borrowed among many languages of south-east Asia due to their central role in religion and cosmology. Words for (100) 'name' are also often borrowed. In fact, if we just look at the English glosses among the items of the 100-word list, we see borrowings for (18) 'person' (from French) (28) 'skin' (from Scandinavian), (32) 'egg' (from Scandinavian), (33) 'grease' (from French) and (86) 'mountain' (from French), among others. Borrowing is a serious problem for the assumption that there is a relatively culture-free basic vocabulary.

Another problem is that glottochronology assumes that there will be a direct, one-to-one matching between each numbered notion in the 100-word list and a word of each language. However, this is very often

not the case. For many of the items on the list, languages often have more than one neutral equivalent. For example, for (1) 'I', many languages of south-east Asia have several forms all meaning 'I' whose use depends on the relative status of the person spoken to. Similarly, (2) 'you' even more frequently than 'I' has multiple forms, depending on social status and degree of intimacy (for example, the familiar versus polite pronouns, Spanish *tu* and *usted*, German *du* and *Sie*, French *tu* and *vous*, Finnish *sinä* and *te*, K'iche' *at* and *la:l*, to mention just a few), where one form is not more basic than the other. For (3) 'we', many languages have distinct forms for 'inclusive' versus 'exclusive' first person plural pronouns. For (8), some languages have no single form for 'not', but rather have conjugated negative verbs with several forms (compare Finnish *en* 'I don't', *et* 'you don't', *ei* 'he/she/it doesn't', *emme* 'we don't', *ette* 'you [plural] don't', *eivät* 'they don't'). For (9) 'all', some languages have different terms depending on whether the meaning is 'all' = 'each member of a group' or 'all' = 'the entire amount'. Navajo and its close sister languages have no unique word for (75) 'water'; rather, they have several different words for 'stagnant water in a pool', 'rain water', 'drinking water' and so on. Some Slavic languages have no unique word for (80) 'cloud', but rather one word for dark storm clouds (as Russian *tuča*) and a separate word for light clouds (as Russian *oblako*). For (84) 'burn', many languages have more than one equivalent; for example, Spanish *arder* 'burn' (intransitive) and *quemar* 'burn (transitive), or several K'ichean (Mayan) languages *-k'at* 'burn' (accidental) and *-por* 'burn' (purposeful). For (93) 'hot', several K'ichean languages have two equally common forms which are equivalent: *k'atan* 'hot' (of weather, water, a room and so on) and *meq'en* 'hot' (of food, drinks, fire and so on). The same is true for (94) 'cold': *te:w* 'cold' (of weather, wind, people, ice and so on) and *xoron* 'cold' (of food, water and so on). K'ichean languages often have as many as seven different terms for 'to eat'; for example, *-waʔ* 'eat (bread-like things)', *-tix* 'eat (meat)' and *-loʔ* 'eat (fruit-like things)' are equally common and none is more neutral or basic than another. Similar examples can be cited for many of the other words in the list.

Not only do many of the items from the 100-word list have more than one natural, neutral equivalent in many languages, but some have no equivalent at all – or better said, in a number of cases, some languages make no distinction between two separate items on the list. For example, (17) 'man' and (18) 'person' are homonymous in many languages. Many languages do not distinguish (27) 'bark' from (28) 'skin' or (36) 'feather' from (37) 'hair', where 'bark' is just 'tree skin', and 'feather'

HP020367

is just 'bird hair'. Some Latin American Indian languages do not distinguish (26) 'root' from (37) 'hair', where 'root' is equivalent to 'tree hair'. More generally, work on colour universals has shown that, while all languages have an equivalent (more or less) for (90) 'white' (or light) and (91) 'black' (or dark) and most have a term for (87) 'red', it is not at all uncommon for languages to lack basic colour terms for (88) 'green' and (89) 'yellow' (Berlin and Kay 1969).

In instances where a language has more than one equivalent per item on the basic vocabulary list or where the same term covers more than one item on the list, the results can be skewed. For example, two languages will appear less closely related than in fact they are if both have, for example, two equivalents for 'hot', but the one meaning 'hot, of weather' turns up checked/ticked on one language's list and the one meaning 'hot, of food' gets checked/ticked on a related language's list. Similarly, if related languages make no distinction between 'feather' and 'hair', then the same word will turn up twice, as the equivalent to these two separate items in the list, making the languages seem to share more and therefore appear to be more closely related than would be the case if only distinct items were compared. Such skewing is a serious problem for the method.

Some 'basic vocabulary' appears to change rather easily for cultural reasons, for example, terms for (38) 'head' in various languages. Proto-Indo-European **kaput-* 'head' gave Proto-Germanic **haubidam/*haubudam* (hence Old English *hēafod* > *head*) and Proto-Romance **kapūt*. However, several Germanic and Romance languages no longer have cognates of these terms as the basic form referring to the human head. For example, German *Kopf* 'head' originally meant 'bowl'; the cognate from **kaput* is *haupt*, which now means basically only 'main', 'chief', as in *Hauptbahnhof* 'main/central train station'. French *tête* and Italian *testa* both meant originally 'pot'; the French cognate from Latin **kapūt* is *chef*, but this means now 'main, principal, chief', not a human head. The Italian cognate *capo* now means 'top, chief, leader'. Pipil (Uto-Aztecan) *tsuntekumat* 'head' comes from *tsun-* 'top, hair (in compound words only)' + *tekumat* 'bottle gourd', and has replaced Proto-Nahua **k^{w}āyi-* for 'head'. It is a problem for the method that some items on the list seem to be replaced more frequently and more easily than others.

Finally, it has been pointed out that taboo has resulted in the replacement of considerable vocabulary, particularly in some languages in Australia, New Guinea and the Americas, where words similar to the names of recently deceased relatives are avoided and substitutions or circumlocutions are used instead. Some of these result in permanent

vocabulary replacement. Other kinds of taboo replacement of items in the basic vocabulary list are also very frequent. For example, in dialects of K'iche' and Tz'utujil (Mayan languages), (20) *ts'ikin* 'bird' has been replaced by *čikop* (originally 'small animal') due to taboo. In Latin American Spanish, *pájaro* 'bird' has come to mean 'male genitals' and is obscene; for that reason, many Spanish speakers avoid it and substitute *pajarito* 'small bird' or something else instead. Because Spanish is the dominant national language where Mayan languages are spoken, speakers of some Mayan languages have transferred the obscene associated with 'bird' in Spanish to the term for 'bird' in their native language and for that reason replaced the vocabulary item. Another example is (32) 'egg'; Spanish *huevo* 'egg' also means 'testicle' and is obscene, and for that reason many in rural Mexico substitute *blanquillo* (literally 'little white thing') for 'egg', replacing *huevo* in this meaning.

Facts such as these show that there is no universal, culture-free vocabulary for which a one-to-one translation equivalent exists in all languages. Still, stubborn proponents of glottochronology would respond to this criticism that something must account for the portion of the vocabulary which is replaced and it may be borrowing, taboo and so on which bring about that loss.

6.5.3.2 Problems with assumptions (2) and (3)

Since the assumption of a constant rate of retention through time and of a constant rate of loss cross-linguistically are related, criticisms of these two assumptions are considered together.

First, a quick check based on common sense would call these assumptions into question. There are good reasons why sound change might be regular, based on what is known about the structure and limitations of human speech-organ physiology and perception; however, there is nothing inherent in the nature of vocabulary (or in the organization of the lexicon) which would lead us to suspect any sort of regular pattern to lexical change, certainly not that basic vocabulary should be replaced everywhere at the same rate. The study of additional test cases after Swadesh developed the method shows that this doubt based on intuition about vocabulary change is well founded, that there really is no constant rate of loss or retention across languages or through time. Icelandic has retained 97.3 per cent, English 67.8 per cent, Faroese over 90 per cent, Georgian and Armenian about 95 per cent each during the time that these languages have had written attestations. The large difference between Icelandic's 97.3 per cent and English's 67.8 per cent lends little confidence to the claim of an expected 86 per cent, regardless of what

HP020368

the range of error (standard deviation) permitted by the statistical calculation may be. That is, these tests show that the rate is neither constant across time nor the same for all languages.

With respect to the claim of a constant loss through time of 14 per cent for each 1,000-year period, written documentation exists for more than one 1,000-year period for extremely few languages; in Swadesh's thirteen test-case languages, attestations for more than one or two 1,000-year chunks of time are available only for the Coptic and Mandarin cases (the interpretation of which is much less secure). Some scholars argue that it is possible that circumstances were so different in the more remote past that vocabulary loss and retention may have behaved differently in earlier 1,000-year chunks of time from later 1,000-year periods. While this is highly unlikely, without written documentation it is not possible to eliminate the possibility entirely, and it will not do just to assert the constant rate far into the past on the basis of no good evidence.

6.5.3.3 *Problems in calculating dates of separation*

Since the split-ups of language families (or subgroups) are usually not sudden, in principle the notion of attaching a precise date to such gradual diversifications seems overly unrealistic – it is difficult to date a language split. Also, subsequent contact among the sister languages after a split is common; but, as commonly applied, the method makes no effort to distinguish loans that result from such contact from directly inherited cognates. For example, in French and Italian the word for 'head' (Italian *testa*, French *tête* [from earlier *test(a)*]) is similar because French borrowed this form from Italian, which itself had shifted *testa* 'pot' to mean 'head'. That is, in calculating how long ago Italian and French separated from one another, the date is skewed towards a more recent break-up because of this basic vocabulary item which is shared due to contact after they split up.

It is also telling that this basic assumption about being able to calculate the date of separation has been vigorously challenged; or better said, the statistical model upon which glottochronology is based has been severely criticized, although others defend it or try to refine it. The most generous thing that can be said about the mathematical model upon which it is based is that it is controversial.

6.5.4 Purported uses of glottochronology

The principal use to which glottochronology has been put is that of

subgrouping language families. It is sometimes thought that glottochronological calculations of splits provide a fast and easy means for arriving at the internal classification of a language family with no need to undertake the more difficult and time-consuming determination of subgrouping based on shared innovations. However, since glottochronology is unreliable and is discounted by most historical linguists, it should not be thought of as a substitute for the traditional means of subgrouping. It is simply not reliable for this purpose.

On the other hand, some have found glottochronology a useful starting point in beginning to classify large families, such as Austronesian, with a great number of languages (c. 800). Since it would be difficult at the outset to compare all the languages of large families with each other to determine shared innovations among them all, some suggest that a preliminary application of glottochronology can give an idea of the more promising hypotheses which can then later be checked by traditional means. However, it should be recalled that glottochronology used in this way does not find or demonstrate subgrouping relationships, but merely points to directions where other sorts of research may prove fruitful. The other research is still necessary before the groupings can be believed, and such preliminary classifications based on glottochronology may well have to be seriously revised or abandoned.

Some suggest that while the dates offered by glottochronology are not reliable, they nonetheless provide a relative chronology which more or less corresponds with what we know in many actual cases. That is, some scholars who reject glottochronology are still willing to entertain the results as a rough guide to relatively old or relatively young relationships. In the absence of other information which can help to establish linguistic dates, this might seem helpful to some. Still, it must be remembered that many glottochronological dates are known to be inaccurate.

Finally, some have thought that glottochronology might help to establish distant genetic relationships among languages. However, glottochronology cannot find or demonstrate remote relationships; rather, in the application of the method, forms which are phonetically similar in the languages being compared are checked/ticked as possible cognates and then, based on the number counted, a date is calculated for when the languages split up. That is, the method does not find or test distant genetic relationships, but rather just assumes relationship and proceeds to attach a date. This is illegitimate for research on possible remote linguistic relationships.

Glottochronology has given linguistics a bad reputation with some

HP020369

other prehistorians. For example, many archaeologists were initially very happy to embrace its dates, and they frequently proposed interpretations of the prehistory of different peoples and areas which relied on glottochronological dates and attempted to correlate them with other sources of information on prehistory. However, as archaeologists came to find out about the problems of the method and the unreliability of the dates, some felt deceived and some came to believe that linguistics had nothing to offer them. This is unfortunate, for though glottochronology proved misleading, other areas of historical linguistics have an important role to play in the study of prehistory in general (as shown in Chapter 15).

In summary, glottochronology is not accurate; all its basic assumptions have been severely criticized. It should not be accepted; it should be rejected. (For references and discussion, see Campbell 1977: 62–5.) For subgrouping, only shared innovations prove reliable, if the cautions about independently occurring changes and possibly inaccurate reconstructions are kept in mind. The best-defined subgroups are those which are based on a number of shared innovations of the type which are not likely to happen independently or to be diffused across language boundaries.

6.6 Exercises

Return to your reconstructions of Proto-Polynesian (Exercise 5.3), Proto-K'ichean (Exercise 5.7), Proto-Quechuan (Exercise 5.8), and Proto-Saami (Exercise 5.9), and based on your reconstruction and the sound changes that you postulated for each language, attempt to establish the subgrouping in these language families. These subgrouping exercises may prove difficult, depending on what you reconstructed and on the number and kind of sound changes which you postulated in each of these reconstruction exercises.

HP020370

7

Models of Linguistic Change

> It is now an axiom of scientific philology that the real life of language is in many respects more clearly seen and better studied in dialects and colloquial forms of speech than in highly developed literary languages.
>
> (Henry Sweet 1900: 79)

7.1 Introduction

When textbooks on historical linguistics talk about 'models of change', they invariably mean the traditional 'family-tree' model and the 'wave theory', and the conflict that is assumed to exist between them. These are described in this chapter and the conflict between them is reconciled. In particular, the contrasting (but actually complementary) approaches taken by dialectologists and traditional Neogrammarians are examined and clarified, sociolinguistic approaches to language change are brought into the picture, and the related notion of 'lexical diffusion' is put in perspective.

7.2 The Family-tree Model

The family tree (sometimes called *Stammbaum*, its German name) is the traditional model of language diversification. The family-tree model attempts to show how languages diversify and how language families are classified (as described in Chapter 6). A family-tree diagram's purpose is to show how languages which belong to the same language family are related to one another. *Linguistic diversification* refers to how a single ancestor language (a proto-language) develops dialects which in time through the accumulation of changes become distinct languages (sister languages to one another, daughter languages of the proto-language), and

15.3.1.3 Proto-Mixe-Zoquean culture

Mixe-Zoquean is a family of some twenty languages spoken in southern Mexico in the region across the Isthmus of Tehuantepec. It is assumed to have been unified until about 1500 BC, and is considered to be of great cultural significance in the region, since it is argued that bearers of the Olmec archaeological culture (the earliest civilization in the region) were speakers of Mixe-Zoquean languages (see below). The reconstructed vocabulary reveals the following cultural inventory:

Maize complex: corn field, to clear land, to sow, to harvest, seed, maize, to grind corn, leached corn, corncob, corn gruel, to grind grains, to shell corn, lime (used to soften kernels of corn for grinding).
Other cultivated plants (and food plants): chili pepper, bean, tomato, sweet potato, manioc, a tuber (species); chokecherry, custard apple, avocado, sapote, coyol palm, guava, cacao.
Animals and procurement of animal resources: deer, rabbit, coatimundi, honey, bee; fish, crab, to fish with a hook, to fish with a net, canoe.
Religion and ritual: holy, incense, knife/axe (used in sacrifice), to write, to count/divine/adore, to dance, to play music, ceremony, year, twenty, bundle of 400, tobacco, cigar, to smoke tobacco (tobacco was used ceremonially).
Commerce: to sell, to pay, to cost, to buy.
Technology: to spin thread, agave fibre, to twist rope/thread, hammock, cord, water gourd, gourd dish, ladder, house, house pole, adobe wall, rubber, ring, arrow, bed, to plane wood, sandals; remedy-liquor (Campbell and Kaufman 1976; Justeson et al. 1985).

15.3.1.4 Cautions about reconstructed vocabulary

Textbooks are fond of repeating warnings about anachronistic reconstructions, which can complicate cultural interpretation based on the reconstructed vocabulary. For example, Bloomfield, in his reconstruction of Proto-Central Algonquian, found cognates which seemed to support reconstructions for a couple of items which were unknown before contact with Europeans, for example 'whisky'. It turns out that the different languages had created names based on the same compound, 'fire'+ 'water' (for example, Cree *iskote:w-a:poy*, composed of *iskote:w* 'fire' + *a:poy* 'water, liquid'), and this 'firewater' compound found in each of the languages looked like a valid cognate set to support the reconstruction, though it is due either to independent parallel development or to diffusion of a loan translation (calque) among these languages. We have

no secure guarantees against such anachronisms entering our cultural interpretations of the past based on reconstructed vocabulary, although we rely on clues from our knowledge of what things were introduced by Europeans and on the criterion which we will see directly (below) that the age of analyzable terms (ones with multiple morphemes) is not as secure as that of unanalyzable terms (those composed of but a single morpheme). In actual cases, this problem comes up rarely; that is, it is not as serious as it might at first appear to be.

15.3.2 Linguistic homeland and linguistic migration theory

A question which has been of great interest in the study of many language families, and especially of Indo-European, is that of the geographical location of the speakers of the proto-language. Two different techniques have been utilized in attempts to determine where speakers of proto-languages lived, that is, where the linguistic 'homeland' (*Urheimat*) of the family was located. We consider each in turn.

15.3.2.1 Homeland clues in the reconstructed vocabulary

The first technique seeks geographical and ecological clues from the reconstructed vocabulary which are relevant to the location of where the proto-language was spoken, especially clues from reconstructed terms for plants and animals. In this approach, attempts are made to find out what the prehistoric geographical distributions were of plants and animals for which we can successfully reconstruct terms in the proto-language, and then these are plotted on a map. The area where the greatest number of these reconstructible plants' and animals' ranges intersect is taken to be the probable homeland of the language family. We will see how this works in the examples considered below.

For the prehistoric geographical distributions of the plants and animals involved, the information which palaeobotany, biology or other fields can provide is relied on. Due to climatic changes and other factors during the last few thousand years, the range of plants and animals is often not the same today as it was in former times. For example, earlier it was argued, based on the reconstruction of **bherəĝ-* (**bherh₁ĝ-*) 'birch', that the Proto-Indo-European homeland lay north of the 'birch line' (where birches grow) which today runs roughly from Bordeaux (France) to Bucharest (Romania). However, this interpretation failed; the birch has shifted its habitat significantly over time and formerly extended considerably to the south, and furthermore it has always been present in the Caucasus region (Friedrich 1970: 30). That is, to locate

HP020371

the birch's distribution during Proto-Indo-European times, we must rely on the results of palynology (the study of ancient pollens). While the case of the birch's earlier distribution is clear, this can make matters difficult, since palynological information may not yet be available for some of the regions in question. Also, in many cases we may have only the roughest of estimates concerning the time when the proto-language was spoken. It is difficult to correlate the distribution of ancient plants based on palynology and of languages without some idea of the period of time at which their respective distributions are being correlated (Friedrich 1970).

15.3.2.2 Linguistic migration theory

The other technique for getting at linguistic homelands – called *linguistic migration theory* – looks at the classification (subgrouping) of the family and the geographical distribution of the languages, and, relying on a model of maximum diversity and minimal moves, hypothesizes the most likely location of the original homeland. The underlying assumption is that when a language family splits up, it is more likely for the various daughter languages to stay close to where they started out and it is less likely for them to move very far or very frequently. Therefore, turning this process around, if we look at today's geographical distribution of related languages, we can hypothesize how they got to where they are now and where they came from. This procedure deals not with just the geographical spread of the languages of the family, but rather with the distribution of members of subgroups within the family. The highest branches on a family tree (the earliest splits in the family) reflect the greatest age, and therefore the area with the greatest linguistic diversity – that is, with the most representatives of the higher-order subgroups – is likely to be the homeland. This is sometimes called the *centre of gravity* model (after Sapir 1949: 455). Lower-level branches (those which break up later) are also important, because they may allow us to postulate the direction of later migration or spread of members of the family. In this model, we attempt to determine the minimum number of moves which would be required to reverse these migrations or spreads to bring the languages back to the centre of gravity of their closest relatives within their individual subgroups, and then to move the various different subgroups back to the location from which their later distribution can be accounted for with the fewest moves. In this way, by combining the location of maximum diversity and the minimum moves to get languages back to the location of the greatest diversity of their nearest relatives, we hypothesize the location of the homeland.

Let's consider some of the better-known cases in which these two techniques have been employed in order to get a feel for how they work.

15.3.2.3 Proto-Indo-European homeland

There is a very large literature on the question of the Proto-Indo-European homeland (see Mallory 1989, Mallory and Adams 1997: 290–9). While there are a number of competing hypotheses, most mainstream historical linguists favour the view which places the Proto-Indo-European homeland somewhere in the Pontic steppes-Caspian region. The evidence for this comes from a linguistic migration theory, interpretation of geographical and ecological clues in the reconstructed vocabulary of the proto-language, loans and the location of their neighbours from whom they borrowed, and attempted correlations with archaeology (though the archaeological interpretations are subject to dispute).

Proto-Indo-European tree names have been at the centre of some homeland considerations, and Proto-Indo-European **bhāĝo-* (**bheh₂ĝo-*) 'beech' has been given much weight. It was thought that beech did not grow to the east of a line running from Königsberg (in East Prussia) to Odessa (in the Crimea). This would seem to place constraints on the location of the Proto-Indo-European homeland, locating it essentially in Europe. However, there are various difficulties with this. There are doubts about the original meaning of the word; the cognates do not all refer to the same tree; Greek *phēgós* means 'oak' and the Slavic forms mean some sort of 'elder', as for example Russian *buziná* 'elder(berry)'; and no reflexes are known from Asiatic Indo-European languages. If **bhāgo-* did not originally mean 'beech', then arguments based on the distribution of beeches in Proto-Indo-European times would not be relevant. There are phonological problems in that the sounds in the putative cognates for 'beech' in some branches of the family do not correspond as they should. Finally, two species of beech are involved and the eastern or Caucasian beech was (and still is) present in the Caucasus and extended to the east. Therefore, many Indo-European groups would have been familiar with it, not just those of Europe west of the infamous Königsberg–Odessa line (Friedrich 1970: 106–15). The problem with the arguments for the homeland based on this distribution of 'birch' was mentioned already above (15.3.2.1); the current distribution of birches is not the same as it was in Proto-Indo-European times and this nullifies the original argument.

Another important participant in the discussion has been Proto-Indo-European **loḱs-* 'salmon', which was formerly thought to have a limited distribution, involving rivers which flowed into the Baltic Sea – this was seen as indicating a Northern European homeland. However, the original meaning of the word appears to include not only 'salmon' but species of salmon-like trout which are found in a very wide distribution

HP020372

which includes also the Pontic steppes and Caspian region, the current best candidate for the homeland (Mallory 1989: 160–1).

The centre of gravity model, when applied to Indo-European, also suggests this area. (For details of other hypotheses for the Indo-European homeland, see Mallory and Adams 1997: 290–9.)

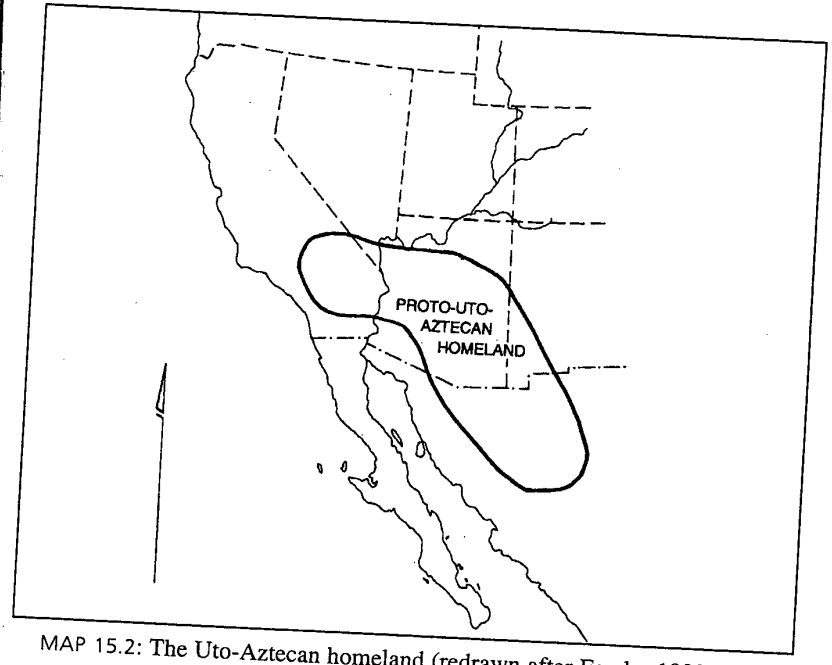
15.3.2.4 Proto-Algonquian homeland

Frank Siebert (1967) found some twenty Proto-Algonquian terms for plants and animals whose distributions overlap in southern Ontario; these animal terms are included among the various ones reconstructed for Proto-Algonquian: golden eagle, pileated woodpecker, oldsquaw, common raven, quail, ruffed grouse, kingfisher, common loon, nighthawk, sawbill duck, seal, raccoon, lynx, squirrel, flying squirrel, moose, porcupine, skunk, fox, bear, woodchuck (groundhog), buffalo (bison), caribou, buck, fawn, beaver, muskrat, weasel, mink, white spruce, tamarack (larch), white ash, conifer–evergreen tree, elm, alder, basswood (linden), sugar maple, beech, willow, quaking aspen; black bass, lake trout, northern pike and brown bullhead. From this he concluded that the original homeland lay between Lake Huron and Georgian Bay and the middle course of the Ottawa River, bounded by Lake Nipissing and the northern shore of Lake Ontario. Dean Snow (1976) reconsidered the Proto-Algonquian homeland focusing on only the names of species whose ranges were most sharply defined; these included five tree names and six animal terms. This resulted in a broader homeland than Siebert had defined, a homeland defined most clearly by the overlap in the territories of the 'beech' and 'tamarack' – the Great Lakes lowlands east of Lake Superior, the St Lawrence valley, New England and Maritime Canada. This was bounded on the west by the Niagara Falls in order to accommodate the reconstructed word for 'harbour seal'. This constitutes a large hunting and trapping zone for nomadic bands. (Considerations mentioned below give a different picture of the Proto-Algonquian homeland.)

15.3.2.5 Proto-Uto-Aztecan homeland

For the Uto-Aztecan family, the results are interesting but not so definitive. Early work on the Proto-Uto-Aztecan homeland had suggested the region between the Gila River and the northern mountains of north-west Mexico, though later work showed that not all the items upon which this conclusion was based could actually be reconstructed in Proto-Uto-Aztecan. Terms which can be reliably reconstructed include, among others, 'pine', 'reed/cane' and 'prickly pear cactus', upon which considerable

attention has been focused. Based on nine certain reconstructions and eighteen less secure but likely reconstructed terms, the Proto-Uto-Aztecan homeland was interpreted to be in 'a mixed woodland/grassland setting, in proximity to montane forests', and this fits a region across south-eastern California, Arizona and north-western Mexico (see Map 15.2) (Fowler 1983).



MAP 15.2: The Uto-Aztecan homeland (redrawn after Fowler 1983: 233)

The results for the Proto-Numic homeland, however, are much more precise. Numic is a subgroup of Uto-Aztecan (to which Shoshoni, Ute and Comanche belong, as well as several others from southern California to Oregon and across the Great Basin into the Great Plains). Catherine Fowler (1972: 119) found that

> The homeland area for Proto-Numic . . . must have been diverse in elevation, allowing for stands of pine and pinyon, but also for such mid- to low-altitude forms as cottonwood, oaks, chia, cholla and tortoises; two, the homeland area was probably in or near desert zones capable of supporting prickly pear, chia, lycium, ephedra, cholla, tortoise, . . . three, based on the presence of proto-forms for cane, crane, heron, mud-hen, tule [reeds], cattail and fish, the area probably contained marshes or some other substantial water sources.

HP020373

She concludes that the Proto-Numic homeland was in Southern California slightly west of Death Valley.

15.3.2.6 Proto-Salishan homeland

Salishan is a family of twenty-three languages spoken on the north-west coast of North America and into the interior as far as Montana and Idaho. From more than 140 reconstructed plant and animal terms in Proto-Salishan, most of which occur throughout the area and thus are of less value in localizing the homeland, M. Dale Kinkade (1991: 143) has determined that some 'two dozen represent species found only on the coast, and hence suggest a coastal, rather than an interior, homeland for the Salish'. These terms include 'harbour seal', 'whale', 'cormorant', 'band-tailed pigeon', 'seagull' (two terms), 'flounder', 'perch', 'smelt' (two terms), 'barnacle', 'horse clam', 'littleneck clam', 'cockle', 'oyster', 'sea cucumber', 'sea urchin', 'red elderberry', 'bracken fern', 'bracken root', 'sword fern', 'wood fern', 'red huckleberry' (two terms), 'salal' (a plant), 'salmonberry' (two terms), 'seaweed', 'red cedar' and 'yew' (Kinkade 1991: 144). Several of these strongly suggest a coastal origin, but not all are equally good as evidence. The terms for 'band-tailed pigeon', 'oyster', 'barnacle', 'sea urchin' and 'flounder' would be supportive, but 'similar forms occur widely throughout the area in several non-Salishan languages and may in the long run turn out to be loanwords; for example, "sea cucumber" and "seaweed" were probably borrowed from neighbouring Wakashan languages' (Kinkade 1991: 147). Proto-Salishan speakers, with their coastal homeland, 'must also have had access to mountains, in particular the Cascade Mountains, because they had names for mountain goats and hoary marmots, both of which are found only at higher elevations' (Kinkade 1991: 147). Based on the distribution of 'bobcats' (not far up the Fraser River) and 'porcupines' and 'lynx' (which did not extend past southern Puget Sound) – for which Proto-Salishan terms are reconstructible – the homeland is further pinpointed:

> extend[ing] from the Fraser River southward at least to the Skagit River and possibly as far south as the Stillaguamish or Skykomish rivers . . . From west to east, their territory would have extended from the Strait of Georgia and Admiralty Inlet to the Cascade Mountains. An arm of the family probably extended up the Fraser River through the Fraser Canyon. (Kinkade 1991: 148)

15.3.2.7 Uralic and Finno-Ugric homeland

Much research has been done on the Proto-Uralic and the Proto-Finno-Ugric homelands, and their identification is on firmer footing than that of Proto-Indo-European (Mallory 2001: 345). These homeland studies often did not distinguish between Proto-Uralic and Proto-Finno-Ugric (a daughter of Proto-Uralic), and many scholars place the homeland of both in the same location. Information from linguistics, archaeology, human genetics and other areas of knowledge has been correlated, generally interpreted in more or less consistent ways, but in hypotheses that differ in their details. For example, the Uralic peoples today have no common culture and are genetically diverse – all Uralic-speaking peoples have received cultural and genetic traits from several directions, in several cases sharing more with non-Uralic neighbours than with other Uralic groups.

Study of the Finno-Ugric homeland has an ample history, though earlier proposals assigning the homeland to central Asia, southern Europe and the like, have now few supporters. The main candidates differ from one another mainly according to the size assumed for the area of the original homeland. They include: (1) the region of the middle course of the Volga River and its tributaries; (2) the region of the northern Urals on both sides of the mountains; (3) the central and southern Urals on both sides; (4) rather eastward on the Asian side of the Urals; (5) rather westward on the European side; (6) the broad area between the Urals and the Baltic Sea. There is actually considerable agreement in these views, since the areas represented are near one another and partially overlapping (Korhonen 1984 and Suhonen 1999; see Map 15.1).

Plant and animal terms have been presented as supporting evidence for hypothesis (1), which is widely held, that the homeland was in the region of the Middle Volga. In view (4), also widely held, the homeland would have been further east and north, between the Urals and the Volga–Kama–Pechora area or on both sides of the Ural Mountains. Supporters of candidate (6) believe that the Proto-Uralic population, at least in its final phases and perhaps also the Proto-Finno-Ugric population, may have occupied a wide area from the Urals to the Baltic Sea, based on the notion that hunting and fishing groups need to exploit wide territories for their subsistence. Ethnographic analogies from subarctic peoples of both the Old and New Worlds have been called upon for supporting evidence, with examples of some reindeer and caribou hunters who travel over 1,000 kilometres twice yearly as they follow the migrating herds of deer (Sammallahti 1984, Mallory 2001). Mikko Korhonen (1984: 63) was of the opinion that while hunting societies

HP020374

typically exploit wide ranges, the proto-language could not have remained unified for long if the speakers were spread from the Ural Mountains to the Baltic. For Korhonen, such a picture could be true, if at all, only briefly at the very end of the unified Finno-Ugric period – the earlier homeland would need to be sought in a smaller area. Pekka Sammallahti (1984: 153), on the other hand, points out that a journey from Lake Ladoga (in the Baltic region) to the Urals (c. 1,200 km) is no longer than from one extreme of Saami territory to the other (c. 1,500 km), and he therefore supposes that a Proto-Uralic or Proto-Finno-Ugric population could have lived in the area between Finland and the Urals and still have maintained a relative linguistic unity (see also Mallory 2001). Hajdú argues that fishing kept the Finno-Ugric people to relatively fixed bases, that 'their manner of life offers no reason for extending their homeland as far as the Baltic' (1975: 38). In any event, most scholars assume that the relative homogeneity of the family was broken up by the introduction of Neolithic techniques and agriculture from areas south of the Proto-Uralic and Proto-Finno-Ugric homeland, that the onset of farming and cattle herding – factors contributing to sedentarism – probably contributed to diversification of the family. Sammallahti points to the uniformity of practically all the paleolithic cultures between the Baltic Sea and the Ural Mountains, which might suggest a linguistic unity, with all the languages of the area perhaps members of a single language family. As long as there were no surplus-producing cultures anywhere nearby, communication among groups was confined to a common ecological (and perhaps cultural) zone, and unity may have been maintained over wide areas by marriage patterns in which spouses as well as linguistic innovations moved from one community to another. However, with the emergence of surplus-producing cultures to the south of the Uralic area, communication was reoriented from latitudinal change to longitudinal change. Longitudinal communication (and weaker latitudinal exchange) caused the ultimate disintegration of the Proto-Uralic area into a series of areas with their own identity and with relatively little interaction, genetic or linguistic, with others. (See also Carpelan 2001.)

Paavo Ravila (1949), employing the techniques of linguistic migration theory, noticed that the Finno-Ugric speaking groups are spread geographically today in a way that reflects their linguistic relationships (degree of relatedness), as though the modern situation was created by movements of these groups to settle in the economically most favourable sections of their former overall territory. Indeed, the region around the middle course of the Volga River with its Oka and Kama tributaries

appears to be a Finno-Ugric centre of gravity; speakers of Mordvin, Mari (Cheremis) and Udmurt (Votyak) live in this region as neighbours, though they represent diverse branches of the family.

Proto-Finno-Ugric vocabulary offers clues for delimiting the homeland; some plant and animal names and some culture words have been considered relevant. The words for 'honey bee' (**mekʃi*) and 'honey' (**meti*) have been emphasized. These were borrowed into Proto-Finno-Ugric from Indo-European. The area where such contact could have taken place is thought to be the region of the middle course of the Volga River, where apiculture was practiced from early times. The honey bee was unknown in Siberia, Turkestan, Central Asia, Mongolia and most of the rest of Asia, but was found in eastern Europe west of the Urals. This area of bee-keeping is often considered one of the clues to the Proto-Finno-Ugric homeland, though this is not without controversy. That the terms refer to wild bees and honey collecting are not ruled out as possibilities (Häkkinen 2001: 176).

There are a sizeable number of reconstructed Proto-Finno-Ugric plant names, but most of these are found in a wide area and are thus not very helpful in limiting the homeland. However, reconstructed tree names have been vigorously discussed in this regard along with five principal trees that have played a role: 'spruce' [Picea obovata], 'Siberian pine' [Pinus sibirica], 'Siberian fir' [Abies sibirica], 'Siberian larch' [Larix sibirica] and 'brittle willow' [Salix fragilis]/'elm' [Ulmus] (outside the Balto-Finnic subgroup the cognates mean 'elm' [Ulmus], compare Finnish *salava* 'willow' and Hungarian *szil* 'elm'). According to Hajdú (1969, 1975), the Finno-Ugric homeland could be located only in an area where all these trees were found at the appropriate time. The only place which fits temporally and geographically is from the Middle Urals toward the north, including the lower and middle course of the Ob and the headwaters of the Pechora river in the area of the northern Urals. Not everyone, however, accepts this interpretation.

Other sorts of vocabulary have also been part of the picture. Cognates for 'hedgehog' have also been taken as evidence for the Finno-Ugric homeland (compare Estonian *siil*, archaic Hungarian *szül-* [syl-] (cf. *sün-* [ʃyn-]); hedgehogs are not found east of the Urals, but do extend as far north as the 61º latitude. A word that has given rise to much speculation is 'metal', with cognates in nearly all Uralic languages meaning 'copper, iron, ore, or metal' (reconstructed as **wäśkä* for Proto-Uralic, seen in Finnish *vaski* 'copper, Hungarian *vas* 'iron', Nenets *veś* 'iron, money', etc.). Since Uralic dates to the Stone Age, such an ancient term for metal is interesting; some suggest the presence of copper trading or

HP020375

cold working of crude copper, but not metallurgy. A metal term of similar shape is also found in various Indo-European (for example, Tocharian A *wäs* 'gold'), and and other languages, so that it may be an old widely borrowed word (Joki 1973: 339–40). It has also been argued that the lack of old terms for 'sea' ('ocean') in Finno-Ugric languages points to a landlocked original homeland (for example, Finnish *meri* 'sea' is a loanword from Baltic Indo-European). There are, however, abundant freshwater terms in the Finno-Ugric vocabulary. Of course, arguments from negative evidence can never be fully persuasive, although this one has been popular. Salminen (2001) believes the reconstructed plant and animal names are not specific enough in their distribution to a conclusion of anything more than that the homeland was far from the sea, in the deep forests rather than in a tundra or steppe environment, though he sees the distribution of the languages as better support for locating the homeland, in the traditional area between the Volga River and the Ural Mountains.

Evidence for the original homeland has also been sought in contacts with other languages. Finno-Ugric has a significant layer of loans from Proto-Indo-European, and also from Indo-Iranian. If we knew the location of Proto-Iranian, perhaps it would help us locate more precisely the Proto-Finno-Ugric homeland. Some scholars argue for even older Indo-European loans in Uralic, though this is controversial, and some others imagine that the loans were all younger but spread across the Finno-Ugric area by diffusion. That is, the testimony of loans is helpful, but apparently not conclusive. Nevertheless, on strong evidence, most scholars believe Proto-Finno-Ugric and Proto-Indo-European were neighbours. (See Joki 1973, Campbell 1997b, Häkkinen 2001, Koivulehto 2001, Sammallahti 2001.)

15.3.2.8 Cautions concerning linguistic homelands migration theory

In linguistic migration theory, the homeland of a language family is inferred to be in the area represented by the greatest diversity (largest number of subgroups) for which the minimum number of moves (migrations) would be required to bring the speakers of the diverse languages back to one place. On the whole, the inferences afforded by this method are strong, and few documented cases fail to conform. In principle, however, it is not difficult to imagine rather straightforward situations in which linguistic migration theory would fail to produce reliable results. For example, suppose a language family with a number of subgroups had once been found in one particular geographical area, but something

forced all their speakers to abandon that area, say a volcanic eruption, a drought, an epidemic or the onslaught of powerful aggressors. In such a case, it is possible that many of the migrating speakers of the different subgroups could end up bunched relatively closely together in a new area, particularly if driven until they encountered some serious obstacle such as insurmountable mountains, an ocean, inhospitable lands without sufficient subsistence resources, or other peoples who prevented entry into their territory. It is also possible that, rather than being driven, several groups speaking languages of the same family might independently be attracted to the same area (or nearby areas), for example to take advantage of better resources available there, to forge alliances with other groups of the area, and so on. In such scenarios, it is in principle possible that we might find that the greatest linguistic diversity would in fact not be in the original homeland, but in the new area where the groups come to be concentrated. Another problem for linguistic migration theory would be the possible situation in which all the languages of a family in the former area of greatest diversity were lost with no trace (where the speakers were annihilated by war or pestilence or whatever), or where the inhabitants remained but their languages were replaced by some other unrelated language or languages. In such a situation, what may appear to be a language family's area of greatest diversity today may not have been that in former times.

The fact that such counter-examples could exist means that the conclusions which we draw from linguistic migration theory can never be absolute, but rather remain inferences, warranted by the evidence but not proven. In our attempts to understand the past, we accept that migration theory has a stronger probability of being correct than any random guess we might make which is not based on these principles. That is, all else being equal, in the absence of other information to help us answer the question, our inference about original homeland based on linguistic migration theory has a better chance of being right than anything else we have to go on.

There are similar problems in relying on clues from reconstructed vocabulary for determining the most likely location of the homeland. One is that groups may migrate to geographical zones where certain flora or fauna of the homeland area are no longer found and as a result lose the words which refer to those items. In such a case, those languages lack the sort of evidence upon which we typically rely to infer the homeland. It is possible that in some cases so many languages have left and as a result lost the relevant vocabulary that these items could not be reconstructed in the proto-language and therefore the evidence for

HP020376

inferring the homeland would be inadequate. To take a specific example, Goddard (1994: 207) finds the terms which Siebert reconstructed 'consistent with the homeland of Proto-Algonquians being somewhere immediately west of Lake Superior' (see above), but points out the circularity of the method. Words for 'harbour seal' would typically only survive in languages in areas where harbour seals are found, leaving out languages (and hence regions) to the west which lacked a cognate for this word. In fact, Goddard concluded that the Proto-Algonquians were located more to the west based on other information, especially the distribution of the languages and the nature of the innovations which they share.

Another problem has to do with instances where the original word is not lost, but its meaning has shifted. Sometimes in such cases it is not sufficiently clear what the proto-meaning may have been to be able to make inferences about the geographical location of its speakers. For example, as mentioned, tree names have played an extremely important role in identifying the Proto-Indo-European and the Proto-Finno-Ugric homelands. If we know what tree names the proto-language had and if we can figure out the geographical distribution of these trees during the time when the proto-language was spoken, we can narrow the homeland down to an area where the distributions of all the trees known in the proto-language intersect. However, semantic shift in some of the tree names to accommodate the fact that the original tree is not found in the new areas to which some groups have migrated, or a shift in the name to accommodate new kinds of trees found in the new areas, severely complicates this sort of research. For example, in Proto-Finno-Ugric, the tree name **sala-* is reconstructed on very solid evidence from across the family; however, as mentioned earlier, this means 'willow' in Finnish and its closer relatives but 'elm' in Hungarian and its closer relatives. That is, we cannot be certain what the testimony of **sala-* is for the location of the homeland of Proto-Finno-Ugric, since the distribution of 'elms' and of 'willows' is quite distinct, but presumably one of these is not the original sense, but rather was acquired as the languages moved out of the territory where the original tree name was known. To take an Indo-European example, even **bherəĝ- (*bherh₁ĝ-)* 'birch', which is one of the best supported of Proto-Indo-European tree names, shifted its meaning to 'ash' in Latin and to 'fir, pine, larch' in Albanian, and is absent in Greek (Friedrich 1970: 29–30; Mallory 1989: 161).

Semantic shifts need not always be a serious problem; in fact, in some cases they can provide us with additional evidence of homeland and migrations away from it. For Proto-Algonquian, a term for

'woodland caribou' is reconstructed based on abundant evidence across many of the branches of the family. This term has shifted its meaning in a few of the languages whose speakers have moved south of the caribou's range. It has come to mean 'bighorn sheep' in the Arapahoan branch and 'deer' in some Eastern Algonquian languages. Because the reconstruction with the meaning 'caribou' is secure on other grounds (distribution across branches of the family), the instances where it has shifted meaning to something else are additional evidence that Arapahoan and those Eastern Algonquian languages involved have moved away from the homeland area where the woodland caribou was found (Goddard 1994).

A problem of a different sort with linguistic homeland models is that they typically imagine a proto-language spoken in a rather restricted region from where groups spread out or migrated to fill up more territory later on. When we go through the exercise of reversing these movements or spreads to the assumed homelands of the various proto-languages, we often find that huge blank areas are left between homelands. The linguistic models seem to imply that these areas were simply not occupied at the time, but typically archaeology finds evidence of human occupation both in the homeland areas and throughout the zones left blank in the linguistic homeland interpretations. These conflicting results need to be accounted for. One possibility is that we have fully misunderstood the nature of how the languages expanded and the territory of the homelands in some cases, though we would like to be able to maintain some faith in these methods. Another possibility is that we do correctly recover the homelands for the most part with our techniques, and that the evidence of human presence in the areas left blank represents languages which have become extinct or been replaced.

15.3.3 Borrowing

Loanwords by their very definition provide evidence of contacts among peoples speaking different languages. The semantic content of loanwords often reveals a great deal about the kinds of contacts that took place and thus about the social relationships among different peoples. The following examples reveal something of the nature and range of historical information that can be retrieved from loanwords in different situations.

A rather straightforward example which illustrates the point about loanwords contributing historical information involves wine-making terms in German, most of which are borrowed from Latin, for example German *Wein* 'wine'< Latin *vīnum*, *Most* 'new wine, must' < *mustum*, *Kelter* 'wine-press' < *calcātūra* 'stamping with the feet' and so on. On

HP020377