

Week 3: Comparative method and linguistic reconstruction

1. Reconstructing a proto-language: By looking at a several genetically related languages we can attempt to reconstruct the ancestor language from which the modern related languages are derived.

2. Some terminology

- *proto-language*: the ancestral language from which modern related (i.e. daughter) languages are descended

Important point: Since we do not have, in most cases, actual access to written records of the proto-language, we work backwards from modern languages or older languages for which we have records to reconstruct the proto-language; *hopefully*, the reconstruction matches reality.

- *cognate*: word (or morpheme) related to a word (or morpheme) in a related (sister) language by virtue of having been descended from a single word (morpheme) of the proto-language
- *cognate set*: the set of related words descending from the same ancestor word of the proto-language
- *comparative method*: method of comparing cognates in order to postulate, or reconstruct, proto-language
- *sound correspondence*: sounds found in the related words of cognate sets which descend from a common ancestral sound
- *reflex*: the descendant sound of a sound in a proto-language

3. How to reconstruct a proto-language

I. Find a cognate set for related languages, or languages suspected to be related

4. It is standard to look for cognates among basic vocabulary items, e.g. body parts, close kinship terms, low numbers, basic geographical terms, since these are more likely to be words which are preserved from the proto-language rather than borrowed at a later time.

5. One common list of words is the Swadesh list (see next page)

6. Here is a small number of cognate sets between Italian, Spanish, Portuguese, and French

Italian	Spanish	Portuguese	French	gloss
kapra	kabra	kabra	ʃεvβ(ə)	goat
karo	karo	karu	ʃεβ	dear
kapo	kabo	kabu	ʃεf	head, top
karne	karne	karne	ʃεβ	meat, flesh
kane	kan	kãw̃	ʃjẽ	dog

7. Establish sound correspondences between languages for the cognate sets

Correspondence 1:

Italian k-: Spanish k-: Portuguese k-: French ʃ-

Note that hyphen after sound indicates that sound is in initial position; hyphen on either side indicates medial sound; hyphen before sound indicates final sound

8. *Important point: Avoid apparent sound correspondences attributed to chance*

9. Example of a false cognate Kakchikel Mayan mes ‘mess, disorder, garbage’ English mess

10. If we look compare other words in the two languages, we see that m:m correspondence is not regular

English	Kakchikel
man	atsi
mouse	tʃʷoj
moon	qatiʔt
mother	nan

II. Reconstruct the proto-sound

11. *How does one reconstruct proto-sounds?*

12. In cases, in which all the daughter languages have the same reflex, it is pretty clear that the reflex found in all the languages will be the sound in the proto-language

13. What about if the reflexes are not identical across languages?

14. Some guiding considerations

15. *Directionality*

16. Reconstruct the sound which allows for formulation of natural (i.e. phonetically) sound changes; naturalness results in greater commonness of change

e.g. given a correspondence s: s: h, we would reconstruct *s, since it is more natural for sounds to lose their oral constrictions than to add an oral constriction

17. In our Romance example, it is more natural for *k* to become *f* than vice versa

18. Similarly, we have a correspondence: Italian -p-: Spanish -b-: Portuguese -b-: French -v-

19. It is natural for stops to become voiced and often spirantized between voiced sounds; this would argue for the changes:

**p* > Spanish *b* / [+voi] __ [+voi]
 > Portuguese *b*
 > French *v*

It would be less natural to have stops devoice in voiced contexts in Italian, i.e. **b* > Italian *p*

20. *Majority wins*

21. Assuming that all languages are equally related, then it is more likely that one language underwent a sound change than multiple languages all independently underwent the same change

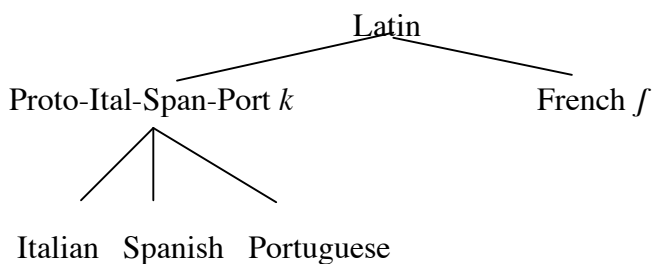
e.g. it is more likely that French changed *k* to *f* than all three other languages independently changed *f* to *k*

22. Important: majority wins is not always reliable

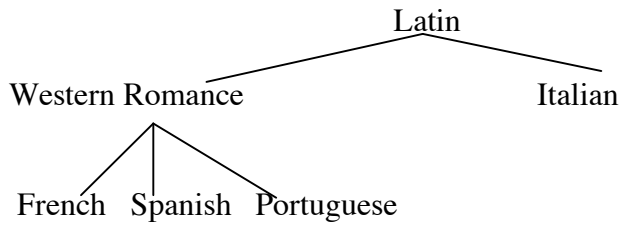
- Majority wins might conflict with naturalness, as in the voicing of **p* discussed above
- A sound change may have occurred independently in separate languages, e.g. diphthongization of high vowels in German and English: German *Haus*, English *house*, Swedish *hus*
- It could be the case that the languages forming the majority form a subgroup and are closely related, in which case the common sound would not be an independent development

e.g. If Italian, Spanish, and Portuguese formed a subgroup of which French were not a member, it could be that the subgroup underwent a change *f* to *k* before the subgroup split up into the daughter languages

In such case, there wouldn't really be a majority, since one sister would have *f* and one would have *k*

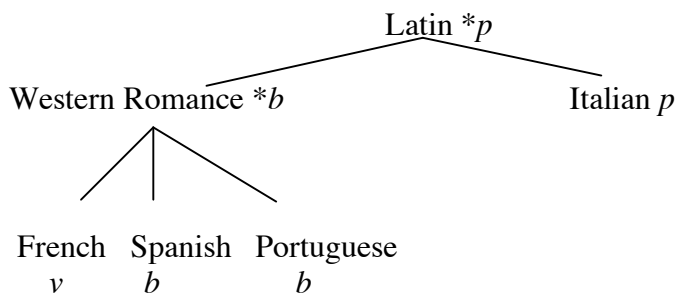


In this particular case, this didn't happen, as the real grouping is as follows



Thus, there is in fact a majority in favor of *k*

Notice that this subgrouping provides a nice account of the facts for **p* discussed earlier



23. *Continuing our investigation of correspondences*

24. It doesn't really matter in what order you proceed: consonants then vowels, first cognate set then second cognate set, initial sounds then medial then final sounds, etc.

25. In general, I prefer to handle the consonants first, since they tend to be more stable over time, but this is just a personal preference

26. So, we could precede an establish a number of correspondences and then reconstruct the proto-sound

Italian a: Spanish a: Portuguese a: French *ɛ* proto **a*
 Italian o: Spanish o: Portuguese u: French *∅* proto **o*

27. In some cases, a proto sound can be associated with more than one correspondence set.

28. Consider the additional cognate sets

Italian	Spanish	Portuguese	French	gloss
kolore	kolor	kor	kulœʁ	color
kostare	kostar	kostar	kute	cost

Italian k: Spanish k: Portuguese k: French k

29. Compare this with our earlier correspondence set
 Italian k-: Spanish k-: Portuguese k-: French f-

30. We have two choices:

- Either the proto-sound for the two correspondence sets is different
- Or there is a difference in environment between the two correspondence sets which explains why *k > f in some cases in French but remained k in other cases

31. First, see if the second scenario can account for the data

32. In fact, we see that the f in French occurs before *a (>e) but the k in French occurs before *u

33. Thus, we can reconstruct *k for both correspondence sets and attribute the *k > f in French to a particular environment

34. Some more data

Italian	Spanish	Portuguese	French	gloss
battere	batir	bater	batɛ	to beat
valle	balje	vale	val	valley

35. Two correspondence sets

Italian b: Spanish b: Portuguese b: French b

Italian v: Spanish b: Portuguese v: French v

36. First, we see whether it is possible to reconstruct *b for both sets, in which case we need a way to predict when we get b in Italian, Portuguese, and French and when we get v.

37. In fact, there is no conditioning factor which allows us to predict when we get b and we get v

38. Thus, we reconstruct different proto-sounds for the two correspondence sets

Italian b: Spanish b: Portuguese b: French b proto *b

Italian v: Spanish b: Portuguese v: French v proto *v

39. There are even cases where the logic of the comparative method dictates that a sound be reconstructed which does not occur in any of the daughter languages

40. The case of Central Algonquian considered by Bloomfield (1925)

Fox	Ojibwa	Plains Cree	Menomini	PCA
hk	sk	sk	tʃk	*tʃk
ʃk	ʃk	sk	sk	*ʃk
hk	hk	sk	hk	*xk
hk	hk	hk	hk	*hk
ʃk	ʃk	hk	hk	*çk

41. The tricky case is the *çk one, but there is little other option since *hk and *jk are both taken for other correspondence sets

42. *Check plausibility of entire reconstructed sound system from typological/naturalness perspective*

43. Suppose we have a correspondence set

Lg. A p: Lg. B pf: Lg. C p: Lg. D pf

44. All else being equal, it might be reasonable to reconstruct either *p or *pf given this correspondence set.

45. We might look at the entire reconstructed system and see if typological plausibility of the sound system might help guide our choice.

46. Suppose we have the following inventory, setting aside the bilabial obstruents

*t *k
*m *n *ŋ
*s
*l

47. From a typological standpoint, it would be more reasonable to reconstruct a proto-language with a *p since *modern* languages without p are pretty rare, but those without *pf are really common.

48. The assumption here is that proto-languages were real languages too and should display similar properties as modern languages.

49. *Important note: Typological plausibility should not be the main factor guiding your reconstruction decisions—it is more crucial to uphold the principles of the comparative method*

50. As a final step, once we have our correspondence sets and their reconstructed sounds we can reconstruct words/morphemes for the proto-language

Italian	Spanish	Portuguese	French	Proto-Romance	gloss
kapra	kabra	kabra	ʃεvβ(ə)	*kapra	goat
karo	karo	karu	ʃεβ	*karo	dear
kapo	kabo	kapu	ʃεf	*kapo	head, top

51. *Glottochronology*: method for dating the split-up of a parent language into daughter languages

52. *Lexicostatistics*: statistical examination of lexical material in order to draw inferences about language history

53. Using lexicostatistics as a technique for examining glottochronology

54. Goal: Look at vocabulary lists and try to figure out the percentage of words that are inherited from the proto-language as opposed to borrowed later on.

55. Get list of basic vocabulary, i.e. Swadesh list, for languages hypothesized/known to be related

56. Separate words into those which are shared cognates and those which are not

57. By examining percentage of shared words, it is hypothesized that we can estimate the time for which the examined languages have been separate, i.e. the time depth of separation from the proto-language

58. It is important that vocabulary be basic, i.e. body parts, numerals, pronouns, geographical terms, basic actions, basic states since non-basic, i.e. peripheral, vocabulary is more likely to be borrowed rather than inherited from the proto-language.

59. Some controversy about whether 100 word list, 200 word list, or 1000 word list is best: balance between increasing statistical robustness of results, amount of material available, and how core, i.e. how likely to not be borrowed, the lexical items are

60. Morris Swadesh, the inventor of glottochronology, tinkered with lists of varying length

61. English: approximately 50% of total vocabulary is borrowed but only about 6% of basic vocabulary

62. *Assumption*: rate of retention of items is relatively constant *for all languages* throughout time: about 80% of the basic vocabulary of 200 items (86% of the 100 item list) is retained over 1000 years and 20% (14% for 100 item list) lost/shifted during this time

63. The rate of retention figures can be used to calculate (as a logarithmic function) the date of divergence between two related languages

64. Criticisms of glottochronology

65. *Notion of basic vocabulary*

66. Is it possible to really come up with a culture-free list of universal “basic” vocabulary items which are not likely to be borrowed?

67. Many basic items not so basic

68. In Mayan lgs., native word for “person” co-existed with a borrowed word for those converted to Christianity

Once everyone was converted, word for “Christian” was generalized to mean person; thus, no need for original word for “person”; the result is borrowing of the word for “person”, a “basic” word

69. Similarly, Spanish word for “dog” borrowed by many Central American lgs., since dogs referred to by native word for did not resemble dogs introduced by European; over time, new word for dog replaced original word

70. Also, often not only a single neutral word for each item on list: e.g. certain languages have multiple words for “I”, “we”, “all”; even “water”, “cloud”, “hot”, “cold”, etc.

Depending on which word is chosen, the retention rate will differ

71. Similarly, two items may share the same translation in a single language, e.g. “man” and “person” might be the same, or “bark” and “skin”, or “feather” and “hair”, etc.

Duplicate items on list will boost the retention rate in a misleading manner

72. Constant rate of retention

73. Swadesh’s original study heavily biased toward Indo-European; 11 of 13 comparisons

74. It turns out, that retention rate for vocabulary varies from language to language even with Indo-European: Iceland 97.3% over last 1000 years, English 67.8%, Georgian and Armenian 95%