Understanding Language Evolution Using an Event-Based Model

David Goldstein
Department of Linguistics
University of California, Los Angeles
dgoldstein@humnet.ucla.edu

Introduction

Modern languages are related to one another through a complicated history of divergence and word borrowing. The divergence of languages is caused by the slow change in spoken language as it is passed from parents to offspring. Over time, divergence causes languages to become increasingly different from one another, ultimately to the point where they are mutually unintelligible. Languages that were spoken by the same human group more recently in time are considered to be more closely related to each other than they are to groups that spoke the language more distantly in time; this relatedness information can be depicted by a tree-like diagram called a 'phylogeny.' Linguistic borrowing, by contrast, causes languages to become more similar to one another.

Language	IPA	Coding
English	/hænd/	0
German	/hant/	0
French	/mẽ/	1
Spanish	/mano-/	1
Italian	/ma:no-/	1
Russian	/rʊka/	2
Polish	/rɛŋka/	2

Table 1. Coding of lexical cognates for the word *hand*.

Languages



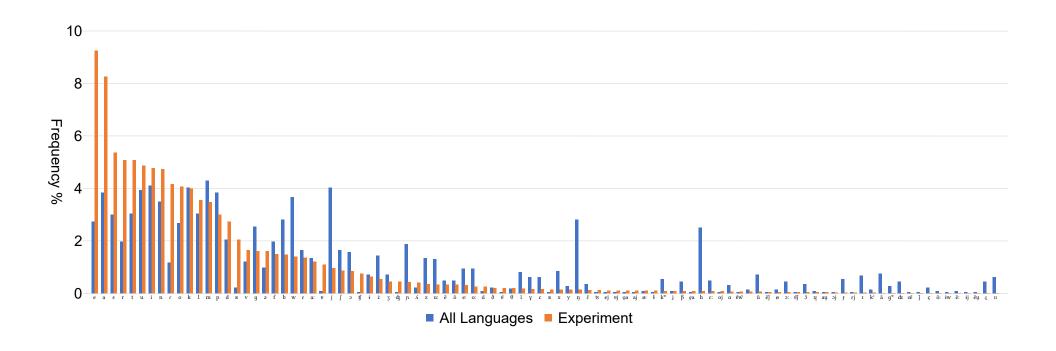
Concept Examples

	I	We	YouPlural	YouSingular	YouSingular- VostramMercedem	Woman
Latin	- e g o:	n or - s	wo: - s	t u:		f e: - m i - n - a
French	3 ə	n u	v u	t y		f a - m
Spanish	j o	n o - s	b o - s	t u	- u s t e d	- e - m - b r - a
Italian	- i - o	n o: i -	v o: i -	t u		f e m m i - n - a
PortugueseBrazil	- e - w	n oj - s	v oj s -	t u	v o s - e -	f e - m e e
Portuguese	- e - w	$ \mathbf{n} \mathbf{n} - \mathbf{j} $	$ \mathbf{v} \mathbf{o} $ – $ \int$	t u	v o s - e -	f e - m j e
Catalan	3 3		$ \mathbf{b} \mathbf{u} - \mathbf{s} $	t u	v o s t e -	f e - m - b r - a
Walloon	dg 1	n o	v o - s	t y		f œ - m
Friulian	j o	n or	v o: - -	t u		$ f \epsilon - m i - n - \epsilon $
Romanian	j e - w	n oj	v oj - -	t u		f e - m e j e
	Woman-Mulier	Woman-Domina	Child	Child-Ninnus	Wife	Husband
	m u l i e r e m	d o m i n a	$i \mid n \mid f \mid a \mid n \mid t \mid e \mid m$		- s p o n s a	m a r i: t u s
French		d a m	$\left oldsymbol{ ilde{lpha}} ight - \left oldsymbol{f} ight \left oldsymbol{ ilde{lpha}} ight - \left - ight - \left - ight $		e - p u - z -	m а в і
	m u x - e r	d o - - n a	$ \mathbf{i} \mathbf{n} \mathbf{f} \mathbf{a} \mathbf{n} \mathbf{t} \mathbf{e} -$	n i n o	e s p o - s a	mariðo-
	m o h h e	d o n - n a	- - f a n t e -	n i n o	- s p o - z a	m a r i t o -
	m u V - 6 R	d o n v	$ ilde{i} - f ilde{v} - f i -$	n e n ẽj̃	i s p o - z a	m a r i d u -
	$m \mid u \mid \Lambda \mid - \mid e \mid \Gamma \mid - \mid - \mid$	d o n v	$\left \widetilde{\mathbf{i}} \right - \left \mathbf{f} \right \widetilde{\mathbf{g}} \left - \left \mathbf{t} \right \widetilde{\mathbf{i}} \right - $	n e n ẽj̃	i f p o - z e	m e r i ð u -
	m u \lambda - e - - -	d o - - n ə		n e n -	$ \mathfrak{d} s \mathfrak{p} \mathfrak{d} - \mathfrak{z} \mathfrak{d}$	m a r i t - -
Walloon		d œ m - - -	$ \epsilon $ - $ \mathbf{f} $ $ \tilde{\mathbf{a}} $ - $ - - $ -			
	m u i: r	d o - - n e				m a r i: t
Romanian	m u - j e r e -	d oam - n ə				m ə r i t
	Husband-Sponsus	Husband-Socius	Mother	Father	Father-Tata	Animal
Latin	- s p o n s u s	s o k i u s	m a: t r e m	p a t r e	t a t a	a n i m a l i a
French	e - p u		m e -	ре-в-		o - m a - j -
Spanish	e s p o - s o -		m a ð r e -	рабге	t a t a	a 1 i m a n - a
Italian	- s p o - z o -	s o tf - o -	m a: d r e -	p a: d r e	t a t a	
PortugueseBrazil	i s p o - z u -		m ej	p aj	t a t a	
Portuguese	$\mathbf{i} \int \mathbf{p} \mathbf{o} - \mathbf{z} \mathbf{u} -$		m ej	p aj	t a t a	
Catalan	ə s p ə - s		ma-rə-	pa-rə		
Walloon			m e - R	p e: - r -		
Friulian	- s p o: - s - -		m a - r i -	p a - r i		- n e m a: 1
Romanian		s o ts			t a t ə	- n ə m a - j e

Character Assignments

Each segment gets a different number

Prior Segment Frequencies



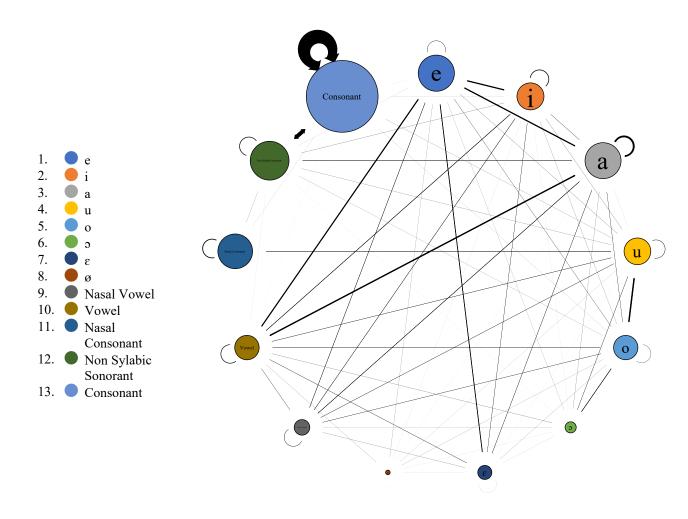
Partition Assignments

Model: "Linguistically Informed"

```
1) e
                                                                                e e:
                                                                               i i:
                                                                                a a:
       u
                                                                                u u:
        0
                                                                                o: o
                                                                                o o:
                                                                                13 3
         Ø
                                                                                ø: ø
        Nasal Vowel
                                                                               \tilde{\alpha} \tilde{i} \tilde{e} \tilde{e} \tilde{j} \tilde{e} \tilde{j} \tilde{o} \tilde{e} \tilde{e} \tilde{w} \tilde{e} \tilde{o}: \tilde{a} \tilde{e}: \tilde{o} \tilde{e} \tilde{u} \tilde{u} \tilde{o}
10) Vowel
                                                                                ə ı oj oj y e œ oa i aj ej ej ij iw au ea α εj σ
11) Nasal Consonant
                                                                               nmnŋĵ
12) Non Sylabic Sonorant
                                                                               wjlr
                                                                               g \, \exists \, j \, d\! j \, s \, \int v \, b \, t \, d \, f \, r \, x \, \land \, \kappa \, t\! f \, p \, z \, \check{o} \, k \, t\! s \, R \, \theta \, h \, c \, \beta \, \mathring{f} \, \psi \, d\! z \, g^w \, \varsigma \, k^j \, \mathring{f} \, \gamma \, k^w \, \, \mathcal{I}
13) Consonant
```

Prior Transition Rates

For the 'Linguistically Informed' model, states were grouped into 13 sets. Here, the area of the circles is proportional to the occurance frequencies for each group. The width of the lines is proportional to the rates of transition between the each partition.

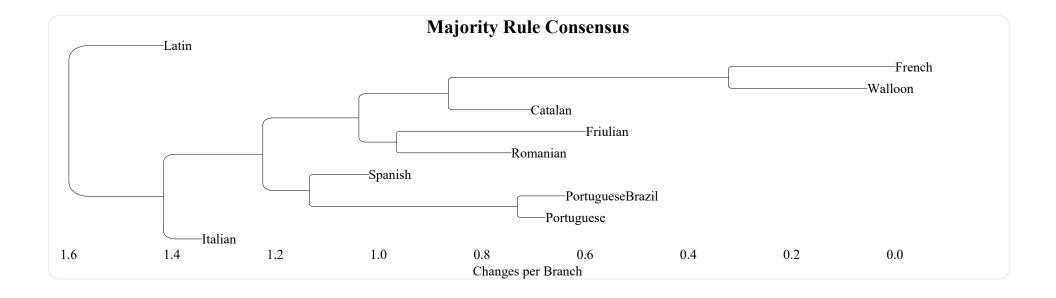


Segment Transition Rates

This chart shows the number of times that each segment (on the left vertical bar) differs from any other other segment in the same cognate set at the same aligned segment position. The rates of transitions to gaps and from gaps are ignored. The segments are grouped into partitions, represented by the sections divided by lines. Within each partition, the segments are ordered from highest to lowest frequency of occurrance. Transition counts are shown with the number in a red box, where the darker reds are larger numbers. The chart has diagonal symmetry.

nignest to lowest frequency of occurrance. Transition counts are shown with the number in a red box, w	nere the darker reas are larger numbers. The chart has diagonal symmetry.
e 112 693 64 631 146 41 8 82 3 29 2 445 19 23 11 8 99 40 17 17 7 11 1 1 325 68 182 33 35 2 13 4 25 10 8 5 4 7	
1 693 26 248 48 1 53 1 43 8 1 99 4 8 1 2 51 22 63 26 2 8 8 1 3 131 6 72 16 13 12 2 1 5 7 7 3 5 4 1	8 6 40 2
a 64 4 248 38 13 2 23 2 1 3 5 2 7 5 1 2 1 3 1 2 1 3 1 2 1 3 1 2 1 3 1 2 1 1 3 1 2 1 1 3 1 2 1 1 1 3 2 2 1 1 3 3 2 2 1 3 3 2 2 1 3 3 2 2 1 3 3 2 2 1 3 3 2 2 1 3 3 2 2 1 3 3 2 2 1 3 3 2 2 1 3 3 2 2 1 3 3 2 2 4 3 3 2 2 4 3 </td <td>- 15 IS - 10 IN IN</td>	- 15 IS - 10 IN IN
64 4 248 38 13 2 2 2 4 15 6 73 12 5 2 4 15 6 73 12 5 2 4 15 6 73 12 5 2 4 4 5 6 73 12 7 7 7 7 7 7 7 7 7	3 3 37 7 4 2 2 7 2
u 8 2 1 1 188 45 2 17 1 1 2 1 1 2 2 12 1 5 3 3 3 1 5 5 a 82 43 29 4 48 45 22 7 6 3 30 3 17 25 15 9	2 1 28 1 6 4 4 2 2 2
or 3 12 35 2 80 29 2 5 2 3 2 4 1 7 2 10 4 1 6 3 2 8 9 3 1 1 3 2 1 7 12 8 2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
x 2 1 34 2 1 2 2 2 4 2 1 1 2 2 2 1 1 1 2 1 2 1	
2 2 5 9 1 4 9 1 12 5 4 0 2 2 2 2 3 4 2 2 2	
22 5 8 1 4 8 8 1 13 5 1 4 9 2 1 1 1 1 8 2 11 2 6 2 2 1 2 3 2 4 2 3 2 4 2 3 2 4 2 3 2 4 3 2 3 2	
2 99 3 51 3 16 2 3	_
2 99 3 51 3 16 2 5 7 9 6 1 1 9 9 9 3 1 1 1 2 3 1 1 1 2 3 1 1 1 1 2 3 1 1 1 1	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	
7 1 8 6 4 1 1 1 5 1 1 1	
11 14 18 4 3 6 2 2 1 1	
1 2	
325 10 131 7 366 8 18 2 22 1 2 23 2 5 9 8 6 3 6 1 1 118 37 2 4 2 1 5 1 1 6 8 2 6 8 2 6 8 16 2 2 7 1 1 21 2 4 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
182 3 72 5 42 1 1 6 1 1 21 1 2 2 9 8 10 4 2 2 1 1 1 1 37 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
33 4 13 7 1 33 4 13 7 1 1 2 2 1 10 1 <td< td=""><td></td></td<>	
1	<u> </u>
0 4 5 9 17 10 7 4 4 4 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	
0 10 2 7 12 3 5 25 4 12 1 3 5 25 4 12 1 3 5 25 4 12 1 3 5 25 4 12 1 3 3 1 3 3 1 3 3 3	
4 5 1 1 1 1	
3 2	36 174 58 9 12 3 7 1 1
	36 1 7 4
	174 1 7 2 58 7 7 2
	9 48 4 2 1 12 4 131 11 2 49 1 1
6 37 1 28 3 6 10 10 10 10 10 10 10 10 10 10 10 10 10	12 2 10 48 99 73 4 4 50 4 20 11 144 4 99 35 1 3 4 99 6 8 10 18 2 185 21 51 1 2 2 6 5 1 3 2 1 51 2 2 7 2 2 7 2
10 40 55 7 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 4 34 41 5 7 325 114 1 3 59 64 12 20 23 8 6 2 2
	12 4 34 3 37 5 152 3 4 2 12 141 14 2 1 69 6 1 11 1 45 1 1 7 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	1 4 10 6 41 37 11 124 86 140 7 4 13 9 11 37 87 13 6 58 1 12 5 1
d	731 11 2 3 522 4 29
V V V V V V V V V V	4 185 30 6 2 32 342 6 6 6 2 28 47 4 74 3 4 3 5 2 9 45 49 57 4 1 3 1 4 11
	21 3 32 9 22 7 1 1 3
	325 12 86 36 22 16 5 15 9 7 1 2 3 1 3
<u> </u>	1 25 1 7 17 6 45 1 9 4 97 7 1 5 13 8 5 1
	11 2 12 3 14 4 51 4 6 49 1 9 1 97 5 3 4 1 2 8 5 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	64 6 11 1 15 30 1 11 1 3 2 1
	2 1 37 2 57 5 4 2
	49
	23 45 6 1 12 3
	1 49 6 2 1 1 1 1 2 2
	35 7 7 1 2 8 5 4 1 5 10 28 3 23
	1 1 6 1 3 2 1 3 1
	2 1 5 3
	2 1 11 1 1 1
	2 3
	•

Results



Questions

