Understanding Language Evolution Using an Event-Based Model

dgoldstein@humnet.ucla.edu

Introduction

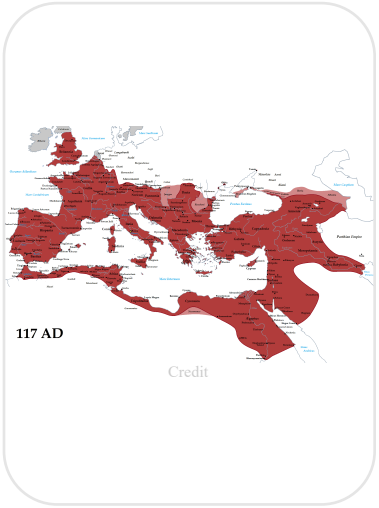
Modern languages are related to one another through a complicated history of divergence and word borrowing. The divergence of languages is caused by the slow change in spoken language as it is passed from parents to offspring. Over time, divergence causes languages to become increasingly different from 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	/maːno/	1
Russian	/rʊka/	2
Polish	/rɛŋka/	2

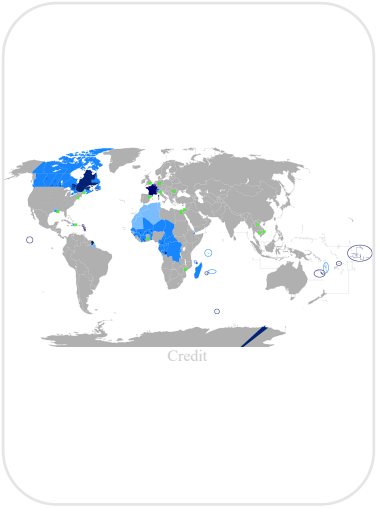
Table 1. Coding of lexical cognates for the word *hand*.

Languages

Latin

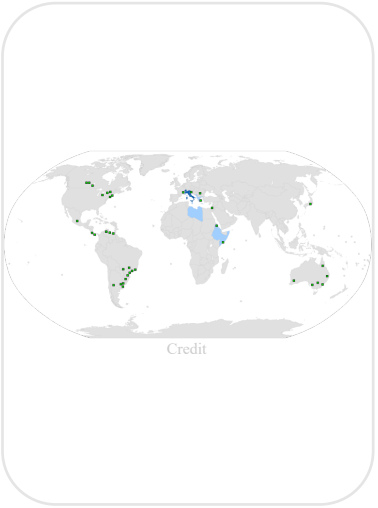


French



Spanish

Italian



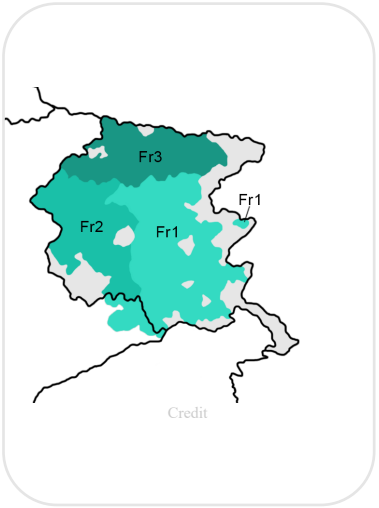
Brazilian Portuguese

Portuguese

Catalan

Walloon

Friulian



Romanian

Concept Examples

	I	We	YouPlural	YouSingular	YouSingular-VostramMerc	Woman
Latin	- e g oː	n oː - s -	w oː - s	t uː		f eː - m i - n - a
French	ʒ ə - -	n u - - -	v u - -	t y		f a - m - - - -
Spanish	ʝ - - o	n o - s -	b o - s	t u	- u s t e d	- e - m - b ɾ - a
Italian	i - - o	n oː i - -	v oː i -	t u		f e m m i - n - a
PortugueseBrazi	- e - w	n ɔ j - s	v ɔ j s	t u	v o s - e -	f e - m e - - - ɐ
Portuguese	- e - w	n ɔ - ʃ -	v o - ʃ	t u	v ɔ s - e -	f e - m j - - - ɐ
Catalan	ʒ - - ɔ		b u - s	t u	v o s t e -	f e - m - b ɾ - a
Walloon	ʤ ɪ - -	n o - - -	v o - s	t y		f œ - m - - - -
Friulian	j - - o	n oː - - -	v oː - -	t u		f ɛ - m i - n - ɛ
Romanian	j e - w	n oj - - -	v oj - -	t u		f e - m e - - j e

	Woman-Mulier	Woman-Domina	ManHumanBeing	ManHumanBeing-Homo	Child	Child-Ninnus
Latin	m u l i e r e m	d o m i n a	h o m i - n e m	h - o m oː	i n f a n t e m	
French		d a m - - -	- ɔ m - - - - -	- - ɔ̃ - -	ɑ̃ - f ɑ̃ - - -	
Spanish	m u x - e ɾ - -	d o - - ɲ a	o m - b ɾ e - -		i n f a n t e -	n i ɲ o
Italian	m o ʎ ʎ e - - -	d ɔ n - n a		- w ɔ m o	- - f a n t e -	n i n o
PortugueseBrazi	m u ʎ - e ʁ - -	d o - - n ɐ	o m ẽj - - - - -		ĩ - f ɐ̃ - ʧ i -	n e n ẽj
Portuguese	m u ʎ - e ɾ - -	d o - - n ɐ	ɔ m ẽj - - - - -		ĩ - f ɐ̃ - t ɨ -	n e n ẽj
Catalan	m u ʎ - e - - -	d ɔ - - n ə	- ɔ m ə - - - -			n e n -
Walloon		d œ m - - -	- ɔ m - - - - -		ɛ - f ɑ̃ - - - -	
Friulian	m u - - iː r - -	d ɔ - - n e	- ɔ m - - - - -			
Romanian	m u - j e r e -	d m - n ə	- o m - - - - -			

	Wife	Wife-Mulier	Husband	Husband-Sponsus	Husband-Socius	Mother
Latin	- s p o n s a	m u l i e r -	m a r iː t u s	- s p o n s u s	s o k i u s	m aː t r e m
French	e - p u - z -		m a ʁ i - - -	e - p u - - - -		m e - ʁ - -
Spanish	e s p o - s a	m u x - e ɾ -	m a ɾ i ð o -	e s p o - s o -		m a ð ɾ e -
Italian	- s p ɔ - z a	m o ʎ ʎ e - -	m a r i t o -	- s p ɔ - z o -	s ɔ ʧ - o -	m aː d r e -
PortugueseBrazi	i s p o - z a	m u ʎ - e ʁ -	m a ɾ i d u -	i s p o - z u -		m ɐ̃j - - - -
Portuguese	i ʃ p o - z ɐ	m u ʎ - e ɾ -	m ɐ ɾ i ð u -	i ʃ p o - z u -		m ɐ̃j - - - -
Catalan	ə s p ɔ - z ə	m u ʎ - e - -	m a ɾ i t - -	ə s p ɔ - s - -		m a - ɾ ə -
Walloon						m e - ʁ - -
Friulian		m u - iː r - -	m a r iː t - -	- s p oː - s - -		m a - r i -
Romanian		m u - j e r e	m ə r i t - -		s o ʦ - - -	

Character Assignments

Each segment gets a different number

1	e	2	g	3	oː	4	ʒ	5	ə	6	ʝ	7	o	8	i	9	w	10	ɔ	11	ʤ	12	ɪ	13	j	14	n	15	s	16	u	17	ʃ	18	oj	19	v
20	b	21	t	22	uː	23	y	24	d	25	f	26	eː	27	m	28	a	29	ɾ	30	ɐ	31	œ	32	ɛ	33	l	34	r	35	x	36	ʎ				
37	ʁ	38	iː	39	ɲ	40	o̯a	41	h	42	ẽj̃	43	ɔ̃	44	ɑ̃	45	ĩ	46	ɐ̃	47	ʧ	48	ɨ	49	p	50	z	51	ð	52	k	53	ʦ				
54	aː	55	ɐ̃j̃	56	ʀ	57	aj	58	ɛː	59	θ	60	ej	61	ɐj	62	õ	63	ẽ	64	β	65	ɬ	66	ɥ	67	øː	68	au̯	69	ø	70	c				
71	g	72	ə	73	ɛj	74	e̯a	75	kʷ	76	ɣ	77	ŋ	78	gʷ	79	ẽu̯	80	ɑ	81	ɟ	82	ɛ̃	83	ɔː	84	ũ	85	ɻ	86	ç				
87	ɑː	88	j̃	89	œ̃	90	ʊ	91	ɐ̃w̃																										

Partition Assignments

Model: “Linguistically Informed”

1 **Nasal Vowel**

ẽj̃ ɔ̃ ɑ̃ ĩ ɐ̃ ɐ̃j̃ õ ẽ ẽu̯ ɛ̃ ũ œ̃ ɐw̃

2 **Vowel**

e oː ə o i ɔ ɪ u oj uː y eː a ɐ œ ɛ iː o̯a ɨ aː aj ɛː ej ɐj øː au̯ ø ə ɛj e̯a ɑ ɔː ɑː ʊ

3 **Nasal Consonant**

n m ɲ ŋ j̃

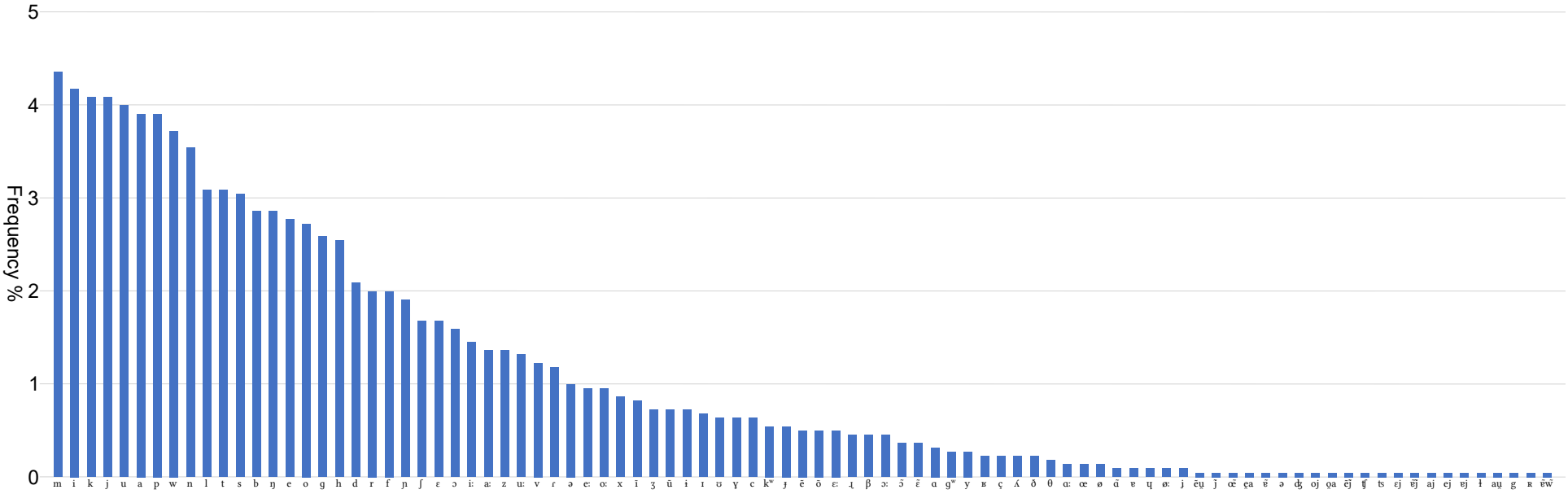
4 **Non Sylabic Sonorant**

w j l r

5 **Consonant**

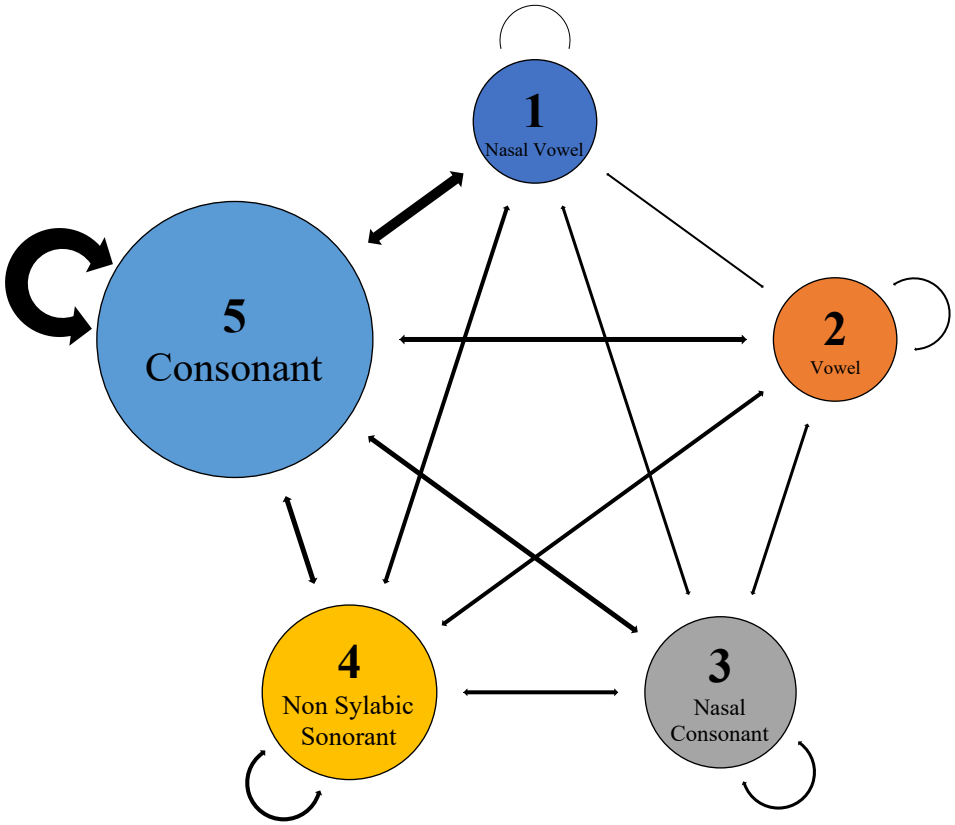
g ʒ ɟ ʤ s ʃ v b t d f ɾ x ʎ ʁ h ʧ p z ð k ʦ ʀ θ β ɬ ɥ c g kʷ ɣ gʷ ɟ ɻ ç

Prior Segment Frequencies

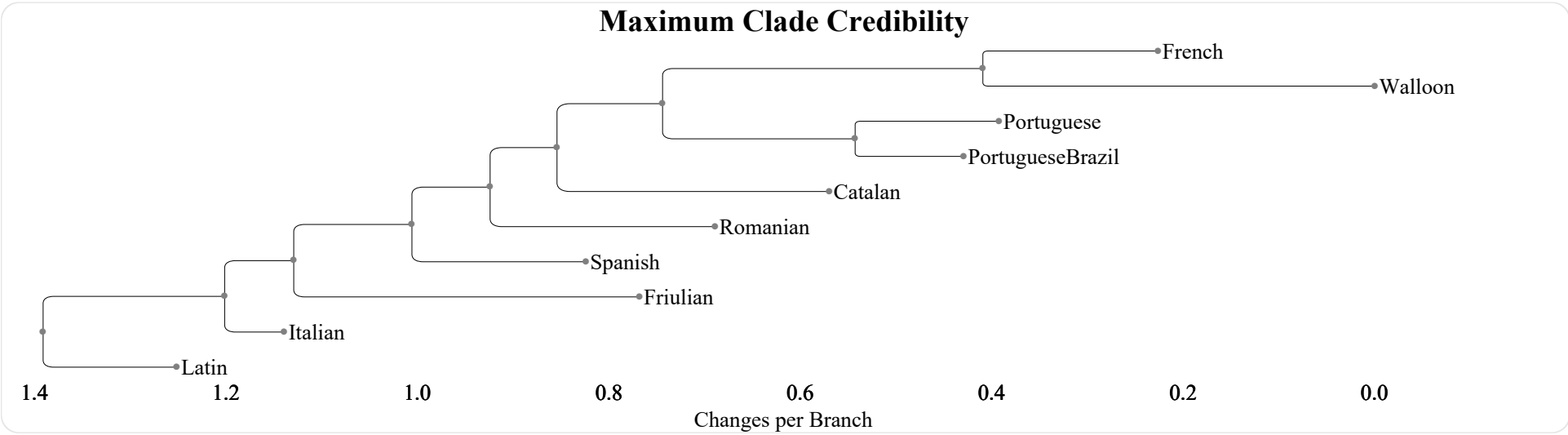


Transition Rates Between Partitions

For the 'Linguistically Informed' model, states were grouped into five sets: Nasal Vowel (1), Vowel (2), Nasal Consonant (3), Non Sylabic Sonorant (4) and Consonant (5). Here, the area of the circles is proportional to the estimated equilibrium frequencies for each group. The width of the arrows is proportional to the estimated rates. Note that rates of change are much greater from one word segment to another when the change is within the word segment group than it is when the change is between word segments in different groups.



Results


Maximum Clade Credibility
French
Walloon
Portuguese
PortugueseBrazil
Catalan
Romanian
Spanish
Friulian
Italian
Latin
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Changes per Branch

Questions