Understanding Language Evolution Using an Event-Based Model

David Goldstein
Department of Linguistics
dgoldstein@humnet.ucla.edu

Introduction

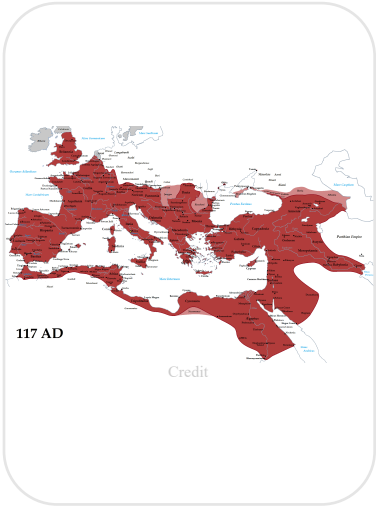
Modern languages are related to one another through a complicated history of divergence and word borrowing. The divergence of languages is caused by the slow change in spoken language as it is passed from parents to offspring. Over time, divergence causes languages to become increasingly different from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

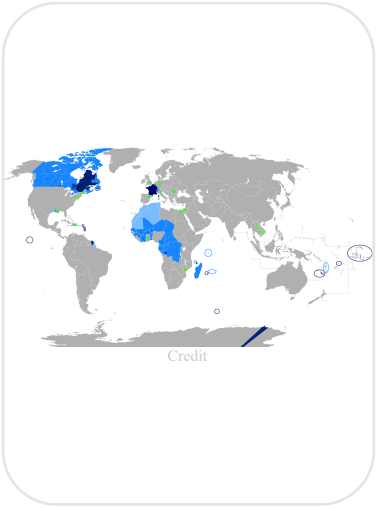
Table 1. Coding of lexical cognates for the word *hand*.

Languages

Latin

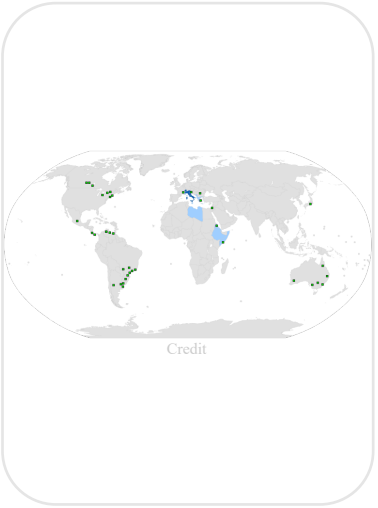


French



Spanish

Italian



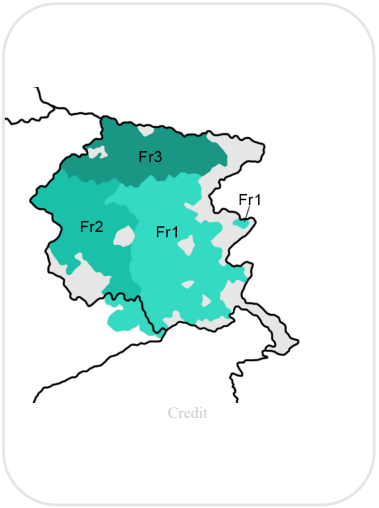
Brazilian Portuguese

Portuguese

Catalan

Walloon

Friulian



Romanian

Concept Examples

	I-Primary	We-Primary	YouPlural-Primary	YouSingular-Primary	YouSingular-VostramMercedem	Woman-Primary
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
PortugueseBrazil	- e - w	n ɔj - s	v ɔj 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	Child-Primary	Child-Ninnus	Wife-Primary	Husband-Primary
Latin	m u l i e r e m	d o m i n a	i n f a n t e m		- s p o n s a	m a r iː t u s
French		d a m - - -	ɑ̃ - f ɑ̃ - - - -		e - p u - z -	m a ʁ i - - -
Spanish	m u x - e ɾ - -	d o - - ɲ a	i n f a n t e -	n i ɲ o	e s p o - s a	m a ɾ i ð o -
Italian	m o ʎ ʎ e - - -	d ɔ n - n a	- - f a n t e -	n i n o	- s p ɔ - z a	m a r i t o -
PortugueseBrazil	m u ʎ - e ʁ - -	d o - - n ɐ	ĩ - f ẽ - ʧ i -	n e n ẽj	i s p o - z a	m a ɾ i d u -
Portuguese	m u ʎ - e ɾ - -	d o - - n ɐ	ĩ - f ẽ - t ɨ -	n e n ẽj	ɨ ʃ p o - z ɐ	m ɐ ɾ i ð u -
Catalan	m u ʎ - e - - -	d ɔ - - n ə		n e n -	ə s p ɔ - z ə	m a ɾ i t - -
Walloon		d a m - - -	ɛ - f ɑ̃ - - - -			
Friulian	m u - - iː r - -	d ɔ - - n e				m a r iː t - -
Romanian	m u - j e r e -	d o̯a m - n ə				m ə r i t - -

	Husband-Sponsus	Husband-Socius	Mother-Primary	Father-Primary	Father-Tata	Animal-Primary
Latin	- s p o n s u s	s o k i u s	m aː 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 - ʁ - -	p e - ʁ -		o - - m a - j -
Spanish	e s p o - s o -		m a ð ɾ e -	p a ð ɾ e	t a t a	a l i m a ɲ - a
Italian	- s p ɔ - z o -	s ɔ ʧ - o -	m aː d r e -	p aː d r e	t a t a	
PortugueseBrazil	i s p o - z u -		m ẽj - - - -	p aj - - -	t a t a	
Portuguese	ɨ ʃ p o - z u -		m ẽj - - - -	p aj - - -	t a t a	
Catalan	ə s p ɔ - s - -		m a - ɾ ə -	p a - ɾ ə		
Walloon			m e - r - -	p ɛ - r -		
Friulian	- s p oː - s - -		m a - r i -	p a - r i		- n e m aː l - -
Romanian		s o ʦ - - -			t a t ə	- n ə m a - j e

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 ɔj 18 ʃ 19 oj 20 v

21 b 22 t 23 uː 24 y 25 d 26 f 27 eː 28 m 29 a 30 ɾ 31 ɐ 32 œ 33 ɛ 34 l 35 r 36 x 37 ʎ 38 ʁ

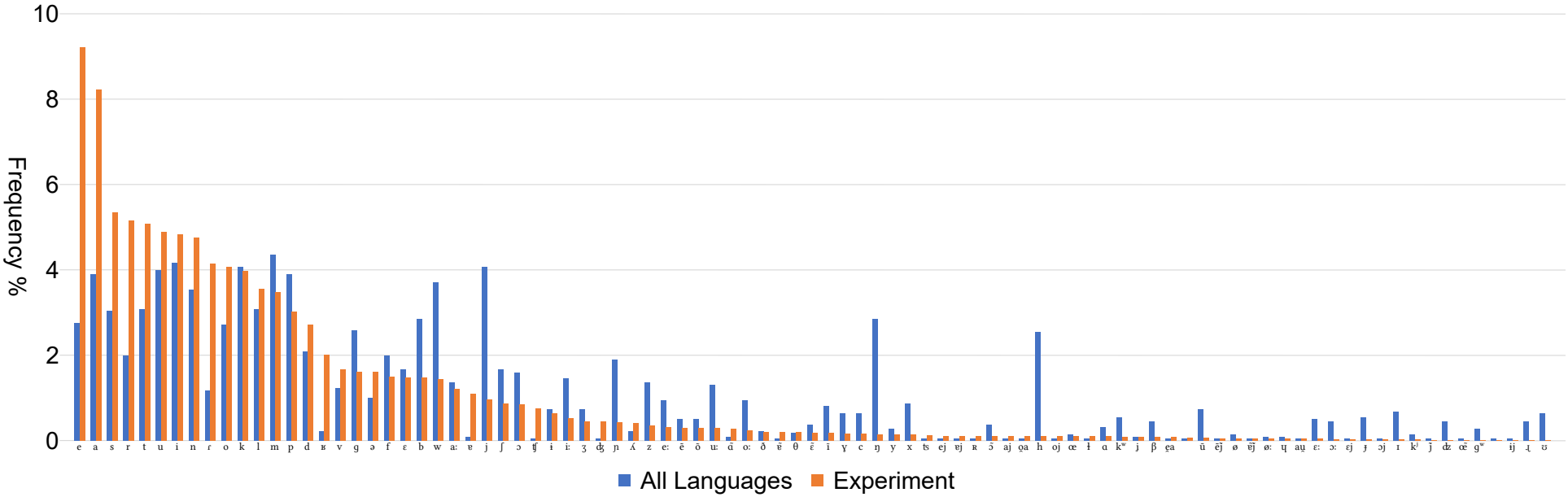
39 iː 40 ɲ 41 o̯a 42 ɑ̃ 43 ĩ 44 ɐ̃ 45 ʧ 46 ɨ 47 ẽj̃ 48 p 49 z 50 ð 51 k 52 ʦ 53 aː 54 ɐ̃j̃ 55 aj 56 θ

57 ej 58 ɐj 59 õ 60 h 61 ɛ̃ 62 ɐ̃w̃ 63 c 64 ɨj 65 ẽ 66 β 67 ɬ 68 ɑ 69 ɥ 70 ɨw 71 ʣ 72 gʷ 73 ŋ

74 ɔ̃ 75 au̯ 76 ø 77 e̯a 78 ʁ 79 ɛj 80 kʲ 81 ɛː 82 ɟ 83 ɣ 84 kʷ 85 øː 86 ɔː 87 ũ 88 ɻ 89 j̃ 90 œ̃

91 ʊ

Prior Segment Frequencies



Partition Assignments

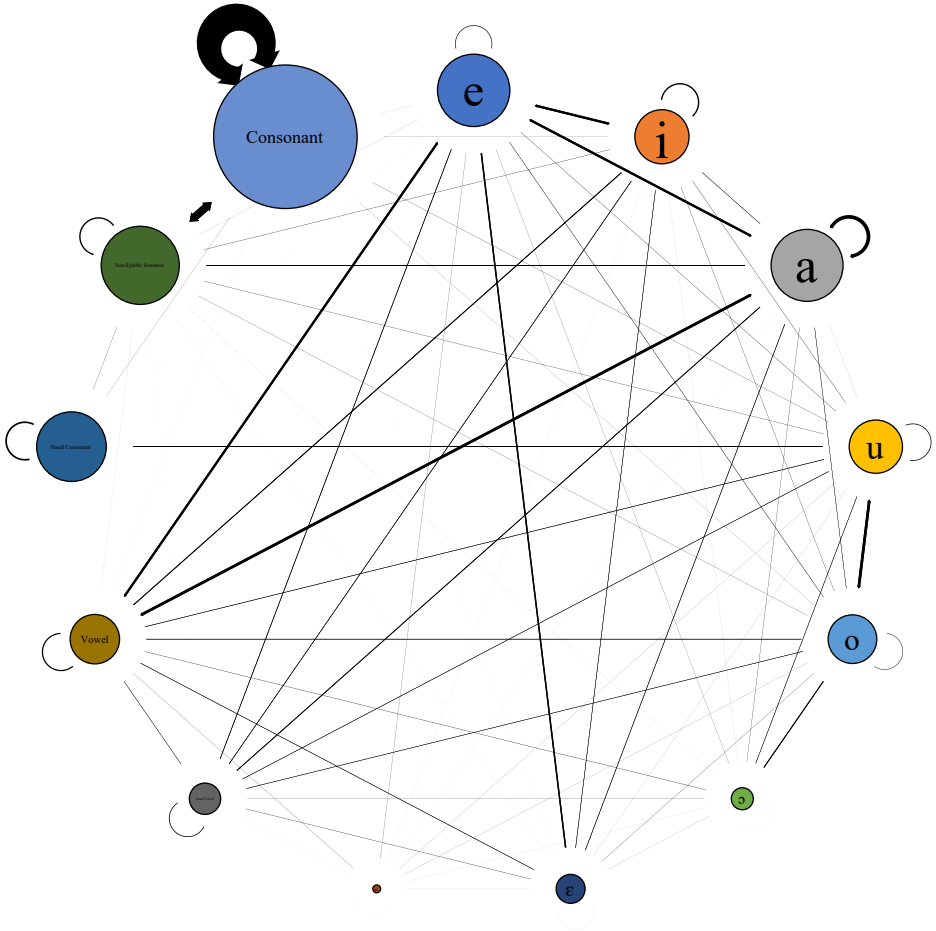
Model: “Linguistically Informed”

1)	e	e eː
2)	i	i iː
3)	a	a aː
4)	u	u uː
5)	o	oː o
6)	ɔ	ɔ ɔː
7)	ɛ	ɛ ɛː
8)	ø	ø øː
9)	Nasal Vowel	ɑ̃ ĩ ɐ̃ ẽj ɐ̃j õ ɛ̃ ɐ̃w̃ ẽ ɔ̃ ũ œ̃
10)	Vowel	ə ɪ ɔj oj y ɐ œ o̯a ɨ aj ej ɐj ɨj ɑ ɨw au̯ e̯a ɛj ʊ
11)	Nasal Consonant	n m ɲ ŋ ɲ̃
12)	Non Sylabic Sonorant	w j l r
13)	Consonant	g ʒ ʝ ʤ s ʃ v b t d f ɾ x ʌ ʁ ʧ p z ð k ʦ θ h c β ɬ ɥ ʣ gʷ ʁ kʲ ɟ ɣ kʷ ɻ

Prior Transition Rates

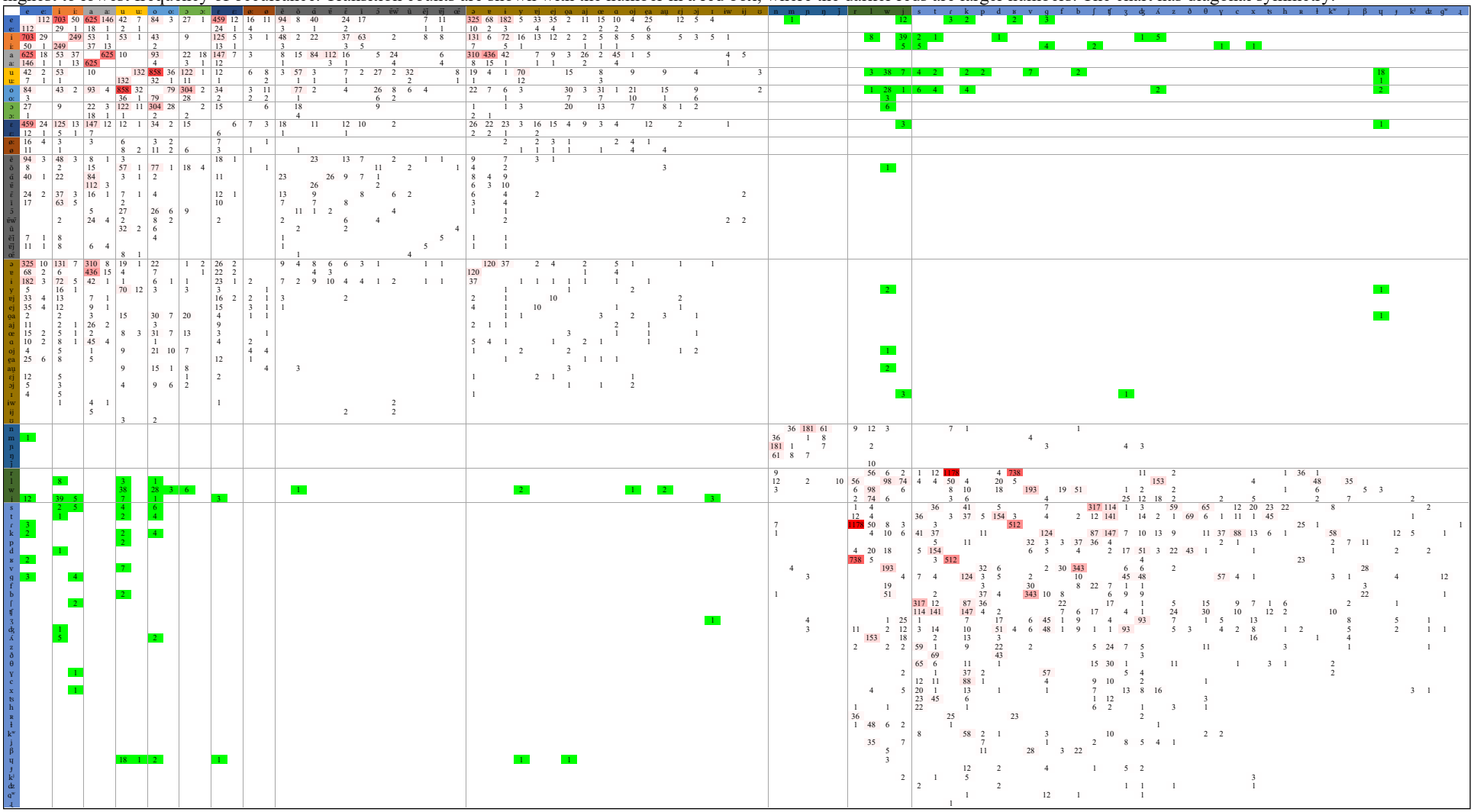
For the ‘Linguistically Informed’ model, states were grouped into 13 sets. Here, the area of the circles is proportional to the occurance frequencies for each group. The width of the lines is proportional to the rates of transition between the each partition.

1. e
2. i
3. a
4. u
5. o
6. ɔ
7. ɛ
8. ø
9. Nasal Vowel
10. Vowel
11. Nasal Consonant
12. Non Sylabic Sonorant
13. Consonant

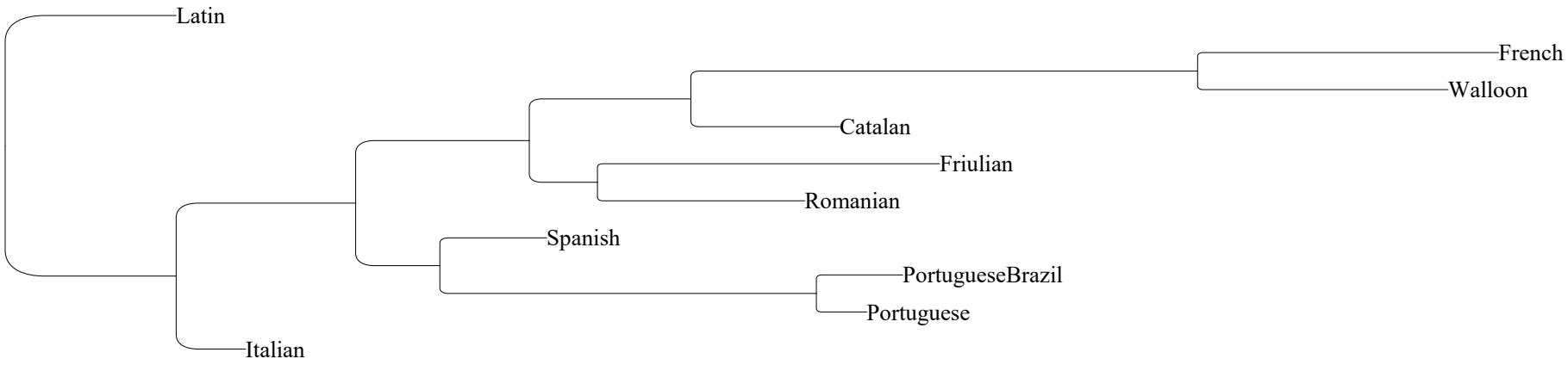


Segment Transition Rates

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.



Results


Latin
French
Walloon
Catalan
Friulian
Romanian
Spanish
PortugueseBrazil
Portuguese
Italian

Questions