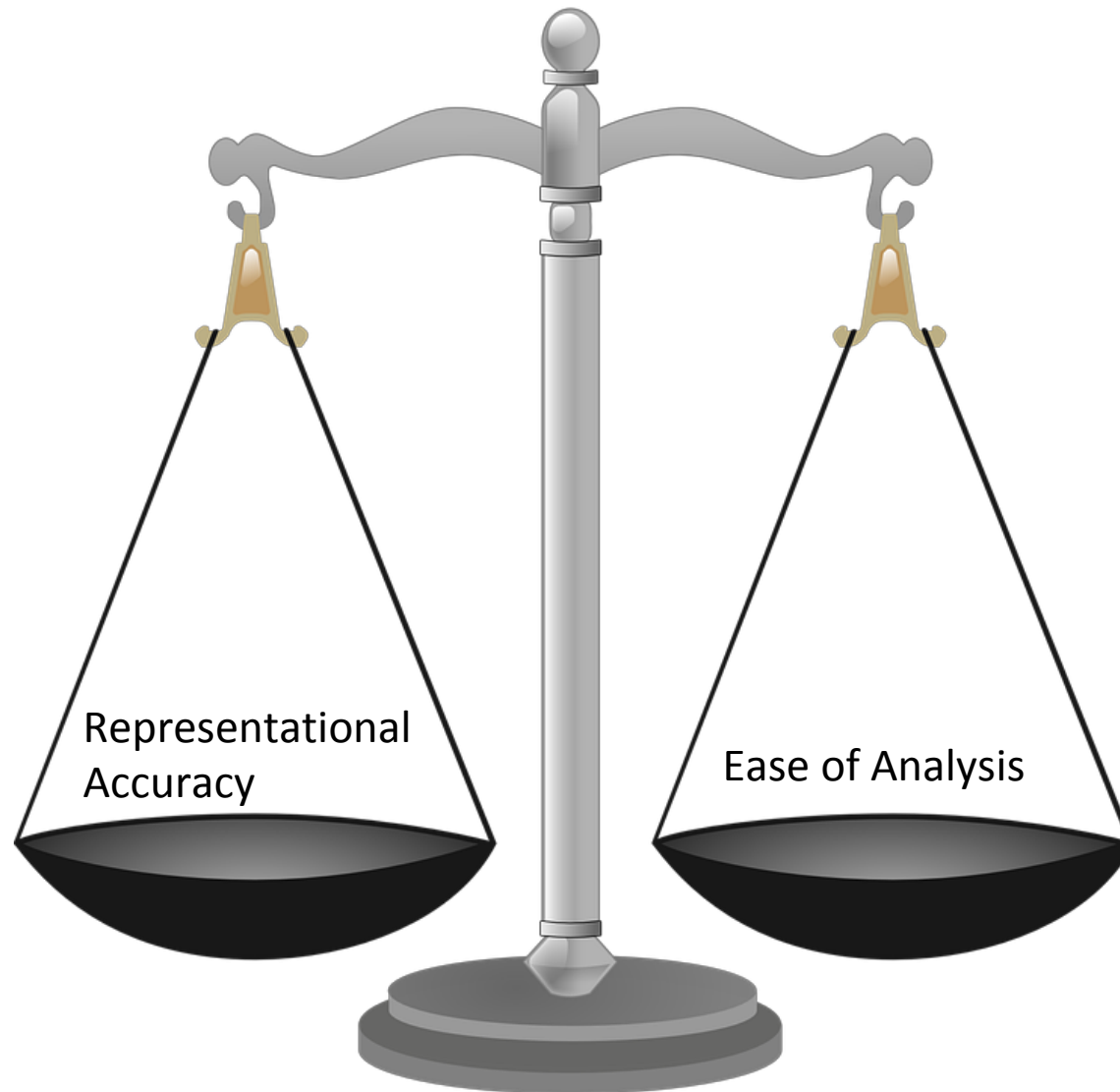


Soundex

- Soundex is a phonetic algorithm used to identify and group words that sound the same.
- Phonetic algorithms help us identify words with different spellings that may be the same or be pronounced similarly, especially names in the US Census.
- Choosing an algorithm for use in social scientific research involves tradeoffs. Choose consciously!

Tradeoffs in phonetic algorithm choice



What kind of harm is lack of representation?

Distributive Harms

How should goods or outcomes be distributed?



Distributive Harms

Equality of Opportunity:
everyone has the same opportunity to pursue good thing



Distributive Harms

**Equality of Outcome:
everyone gets same good things (and same responsibilities)**



Distributive Harms

Equality of Welfare:

everyone has well-being/happiness, but only some people need strings for their happiness



Representational Harms

Am I represented in this system?

Can I express myself in it?

Does this system recognize me, my culture, and my self-expression?

ASCII

“American Standard Code for Information Interchange”
covers 256 symbols

USASCII code chart

Bits					0	0	0	1	1	1	1	1
b ₇	b ₆	b ₅	b ₄	b ₃	0	0	0	1	1	1	1	1
b ₄	b ₃	b ₂	b ₁	Column	0	1	2	3	4	5	6	7
↑	↑	↑	↑	Row ↓	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	\$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(8	H	X	h	x
1	0	0	1	9	HT	EM)	9	I	Y	i	y
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	VT	ESC	+	;	K	[k	{
1	1	0	0	12	FF	FS	,	<	L	\	l	
1	1	0	1	13	CR	GS	-	=	M]	m	}
1	1	1	0	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	SI	US	/	?	O	_	o	DEL

Unicode

Covers over 1 million symbols,
Full coverage of 90 languages,
basic coverage for 200 languages



U+30ED	U+15D2	U+2666	U+043E	U+2690	U+2B07
U+FF0F	U+2B1C	U+1F490	U+2639	U+263B	U+13EA
	<p>Everyone in the world should be able to use their own language on phones and computers.</p> <p>▶ LEARN MORE ABOUT UNICODE</p>				
U+1F643					
U+1F9E1					
U+0994	U+05E9	U+0E07	U+1D25	U+51F8	U+0B37
U+203F	U+1F60A	U+03CE	U+127F	U+060F	U+1557

- Adopt a Character +
- Emoji +
- Basic Info +
- News
- Events
- Connect +
- Membership +
- Press

Unicode and Representation



	☐ U+30ED	➤ U+15D2	◆ U+2666	○ U+043E	⚑ U+2690	↓ U+2B07
	/ U+FF0F	■ U+2B1C	🌸 U+1F490	😞 U+2639	😊 U+263B	℔ U+13EA
Adopt a Character	🙄 U+1F643	Everyone in the world should be able to use their own language on phones and computers.				
Emoji		▶ LEARN MORE ABOUT UNICODE				
Basic Info						
News	❤️ U+1F9E1					
Events						
Connect	ॐ U+0994	ሀ U+05E9	↓ U+0E07	⚓ U+1D25	🏠 U+51F8	⌘ U+0B37
Membership						
Press	◡ U+203F	😊 U+1F60A	ó U+03CE	⚔ U+127F	⋆ U+060F	↷ U+1557

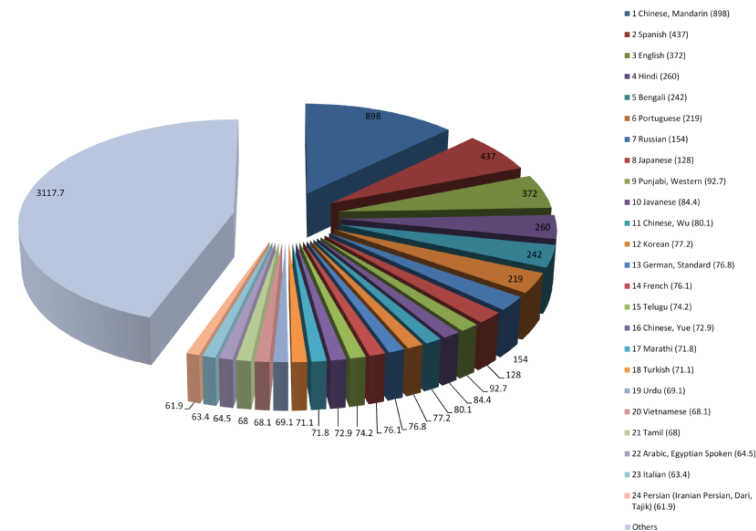
- Adopt a Character +
- Emoji +
- Basic Info +
- News
- Events
- Connect +
- Membership +
- Press

Language Representation in Unicode

There are at least 7,100 languages spoken today.

Of those, there are at least 24 with 50 million speakers.

Are the languages you speak representable in Unicode?



Languages with at least 50 million first-language speakers, millions

Total: 7000 million (2011)

Katie's Office Hours

Come say hello any time:

<https://calendly.com/kathleencreel>