

# Strings & Representation

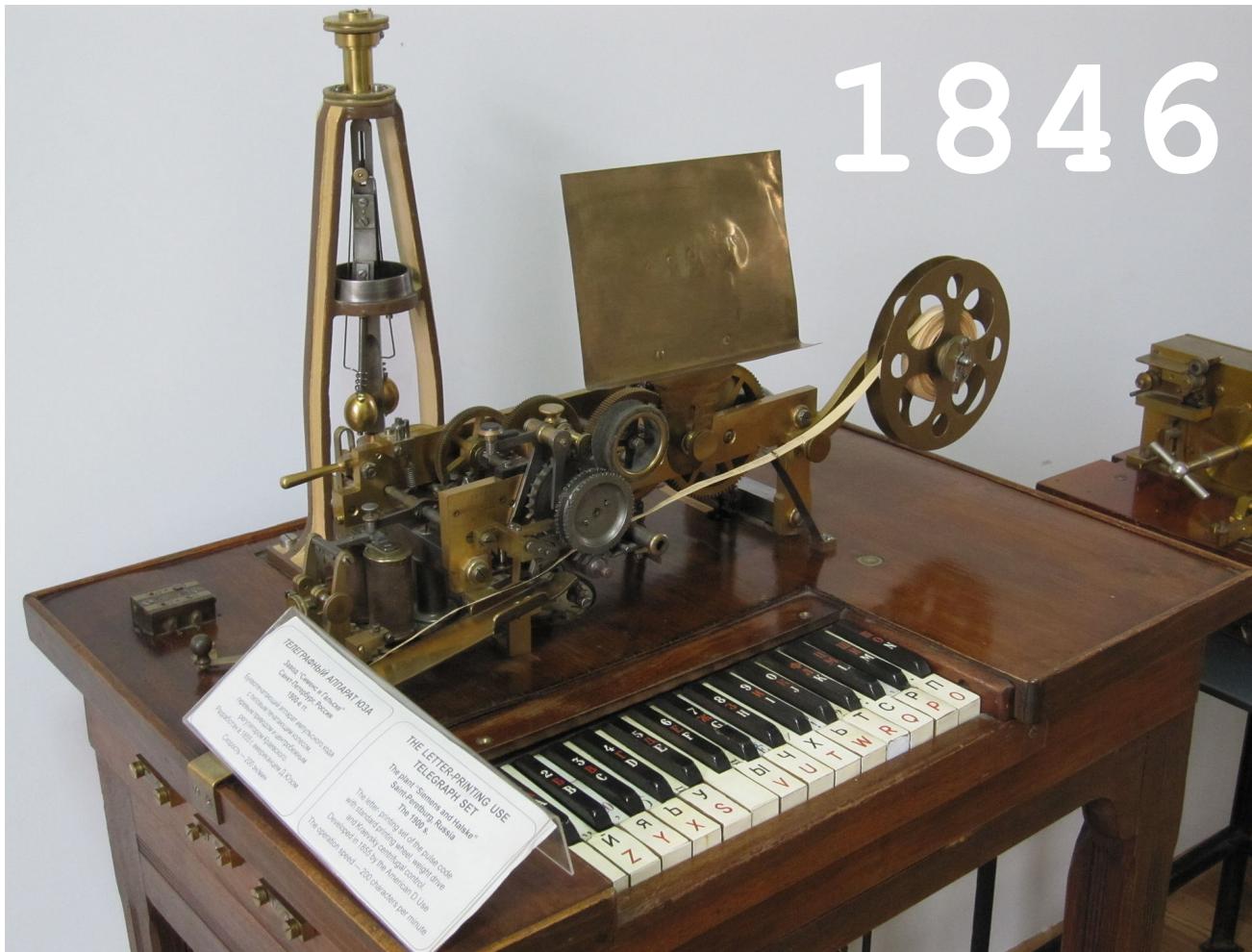
# Learning Goals

## 1. Understanding Representational Harms



# History of Symbol Representation

“Typewheel” for the telegraph,  
covers 56 characters



Piech + Sahami, CS106A, Stanford University



# History of Symbol Representation

# 4-digit character-to-code mapping system for the telegraph

○一八一 𠂔	○六一 讐	○一四一 伏	○一三一 𠂔	○一〇一 仕	○〇九一 了	○〇六一 𠂔	○〇四一 𠂔	○〇二一 𠂔	○〇〇一 𠂔
○一八二 佳	○六二 位	○一四二 伐	○一三二 𠂔	○一〇二 仁	○〇九二 也	○〇六二 𠂔	○〇四二 𠂔	○〇二二 𠂔	○〇〇二 𠂔
○一八三 餅	○六三 低	○一四三 休	○一三三 𠂔	○一〇三 𠂔	○〇九三 子	○〇六三 𠂔	○〇四三 𠂔	○〇二三 𠂔	○〇〇三 𠂔
○一八四 信	○六四 佐	○一四四 伙	○一三四 仔	○一〇四 仆	○〇九四 交	○〇六四 事	○〇四四 丰	○〇二四 丈	○〇〇四 三
○一八五 佾	○六五 佐	○一四五 伯	○一四五 怜	○一〇五 仇	○〇九五 亥	○〇六五 事	○〇四五 乖	○〇二五 串	○〇〇五 三
○一八六 使	○六六 佑	○一六六 佑	○一六六 佑	○一〇六 伸	○〇九六 亦	○〇六六 事	○〇四六 乘	○〇二六 上	○〇〇六 上
○一八七 侃	○六七 佔	○一四七 佚	○一七七 佽	○一〇七 介	○〇九七 享	○〇〇七 二	○〇四七 ○〇二七	○〇二七 下	○〇〇七 不
○一八八 來	○六八 何	○一六八 你	○一七八 𠂔	○一〇八 仍	○〇九八 疵	○〇〇八 疵	○〇〇八 八	○〇〇八 𠂔	○〇〇八 𠂔
○一八九 修	○六九 伐	○一四九 𠂔	○一九九 件	○一〇九 仔	○〇九九 享	○〇〇九 云	○〇〇九 九	○〇〇九 九	○〇〇九 九
○一九〇 例	○七〇 余	○一五〇 伴	○一五〇 件	○一〇〇 仕	○〇九〇 京	○〇〇〇 互	○〇〇〇 乙	○〇〇〇 凡	○〇〇〇 丑
○一九一 待	○七一 余	○一五一 俗	○一五一 俗	○一〇一 攸	○〇九一 亭	○〇〇一 九	○〇〇一 九	○〇〇一 丹	○〇〇一 且
○一九二 侏	○七二 佛	○一五二 伸	○一五二 伸	○一〇二 伎	○〇九二 亮	○〇〇二 井	○〇〇二 乞	○〇〇二 主	○〇〇二 丕
○一九三 𠂔	○七三 作	○一五三 伺	○一五三 伺	○一〇三 任	○〇九三 毫	○〇〇三 亘	○〇〇三 也	○〇〇三 世	○〇〇三 世
○一九四 𠂔	○七四 𠂔	○一五四 𠂔	○一五四 𠂔	○一〇四 彷	○〇九四 亶	○〇〇四 瓦	○〇〇四 𠂔	○〇〇四 丘	○〇〇四 丘
○一九五 𠂔	○七五 𠂔	○一五五 𠂔	○一五五 𠂔	○一〇五 企	○〇九五 亹	○〇〇五 𠂔	○〇〇五 𠂔	○〇〇五 𠂔	○〇〇五 𠂔
○一九六 𠂔	○七六 佩	○一五六 伽	○一五六 伽	○一〇六 𠂔	○〇九六 𠂔	○〇〇六 些	○〇〇六 乾	○〇〇六 父	○〇〇六 垂
○一九七 𠂔	○七七 𠂔	○一五七 𠂔	○一五七 𠂔	○一〇七 𠂔	○〇九七 𠂔	○〇〇七 𠂔	○〇〇七 𠂔	○〇〇七 𠂔	○〇〇七 𠂔
○一九八 供	○七八 𠂔	○一五八 𠂔	○一五八 𠂔	○一〇八 𠂔	○〇九八 𠂔	○〇〇八 𠂔	○〇〇八 𠂔	○〇〇八 𠂔	○〇〇八 𠂔
○一九九 依	○七八 𠂔	○一五九 𠂔	○一五九 𠂔	○一〇九 伍	○〇九九 𠂔	○〇〇九 𠂔	○〇〇九 𠂔	○〇〇九 𠂔	○〇〇九 𠂔
○二〇〇 仰	○八〇 𠂔	○一六〇 𠂔	○一六〇 𠂔	○一〇〇 伎	○〇九〇 𠂔	○〇〇九〇 人	○〇〇九〇 𠂔	○〇〇九〇 𠂔	○〇〇九〇 𠂔



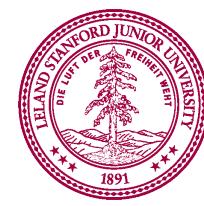
# ASCII

“American Standard Code for Information Interchange” covers 256 symbols

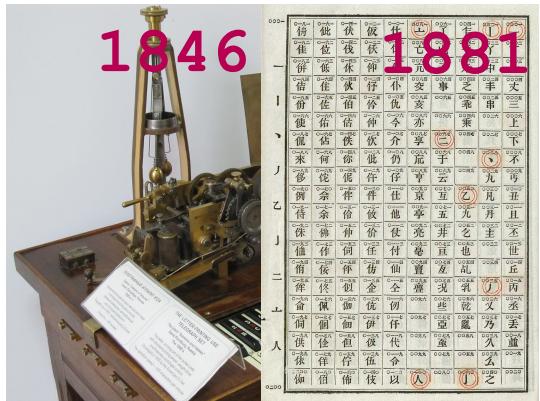
USASCII code chart



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



# History of Character Representation



USASCII code chart

		0	0	0	1	0	1	0	0	1	1	0	1	1	1	1
		b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	b <sub>4</sub>	b <sub>5</sub>	Column	Row	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	NUL	DLE	SP	0	@	P	\	p	
0	0	0	0	1	1	1	1	SOH	DC1	!	1	A	Q	a	q	
0	0	1	0	2	2	2	2	STX	DC2	"	2	B	R	b	r	
0	0	1	1	3	3	3	3	ETX	DC3	#	3	C	S	c	s	
0	1	0	0	4	4	4	4	EOT	DC4	\$	4	D	T	d	t	
0	1	0	1	5	5	5	5	NAK	ENQ	%	5	E	U	e	u	
0	1	1	0	6	6	6	6	ACK	SYN	B	6	F	V	f	v	
0	1	1	1	7	7	7	7	BEL	ETB	'	7	G	W	g	w	
1	0	0	0	8	8	8	8	BS	CAN	(	8	H	X	h	x	
1	0	0	1	9	9	9	9	HT	EM	)	9	I	Y	i	y	
1	0	1	0	10	10	10	10	LF	SUB	*	:	J	Z	j	z	
1	0	1	1	11	11	11	11	VT	ESC	+	:	K	L	k	(	
1	1	0	0	12	FF	FS	.	<	L	\	l	/				
1	1	0	1	13	CR	GS	-	=	M	J	m	)				
1	1	1	0	14	SO	RS	.	>	N	\^	n	\~				
1	1	1	1	15	S1	US	/	?	0	—	o	DEL				

1963



Adopt a Character

Emoji

Basic Info

News

Events

Connect

Membership

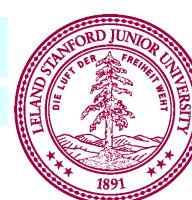
Press

1991



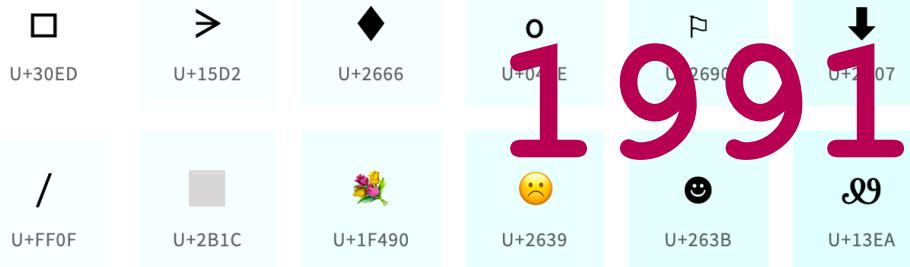
Everyone in the world should be able to use their own language on phones and computers.

LEARN MORE ABOUT UNICODE



# Unicode

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



Adopt a Character



Everyone in the world should be able to use their own language on phones and computers.

Emoji



U+1F643

Basic Info



[LEARN MORE ABOUT UNICODE](#)

News

U+1F9E1

Events

Connect



Membership



U+0994

U+05E9

U+0E07

U+1D25

U+51F8

U+0B37

Press

U+203F

U+1F60A

U+03CE

U+127F

U+060F

U+1557



# Representing a Burrito in Unicode/UTF-8

Emoji added in 2007

Burrito (U+1F32F) added in 2015

---

Apple



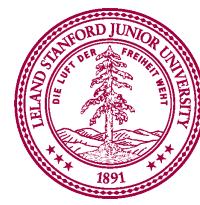
---

Google



---

Microsoft



# Unicode and Representation



□ U+30ED > U+15D2 ♦ U+2666 o U+043E ♡ U+2B07

/ U+FF0F > U+2B1C ♡ U+1F490 o U+2639 ♡ U+263B ♡ U+13EA

Everyone in the world should be able to use their own language on phones and computers.

[LEARN MORE ABOUT UNICODE](#)

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

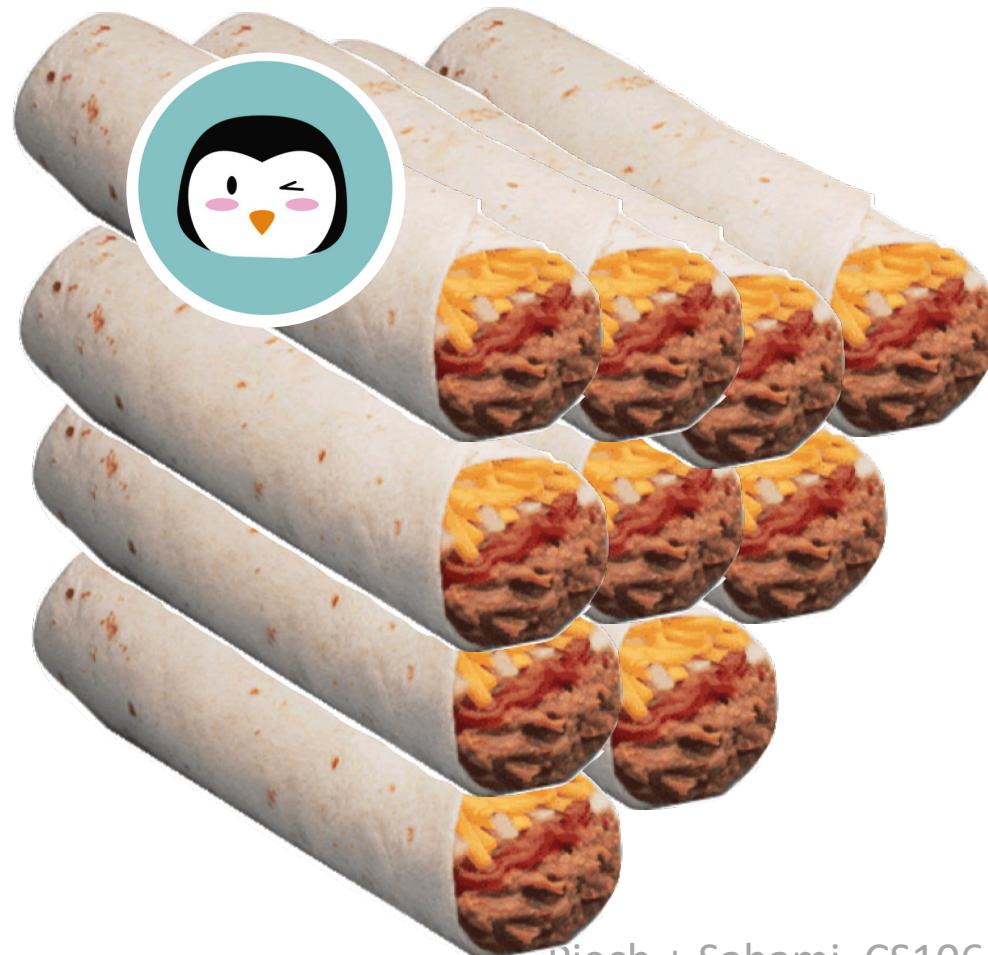
□ U+30ED > U+15D2 ♦ U+2666 o U+043E ♡ U+2B07

□ U+30ED > U+15D2 ♦ U+2666 o U+043E ♡ U+2B07

What kind of harm is a  
lack of representation?

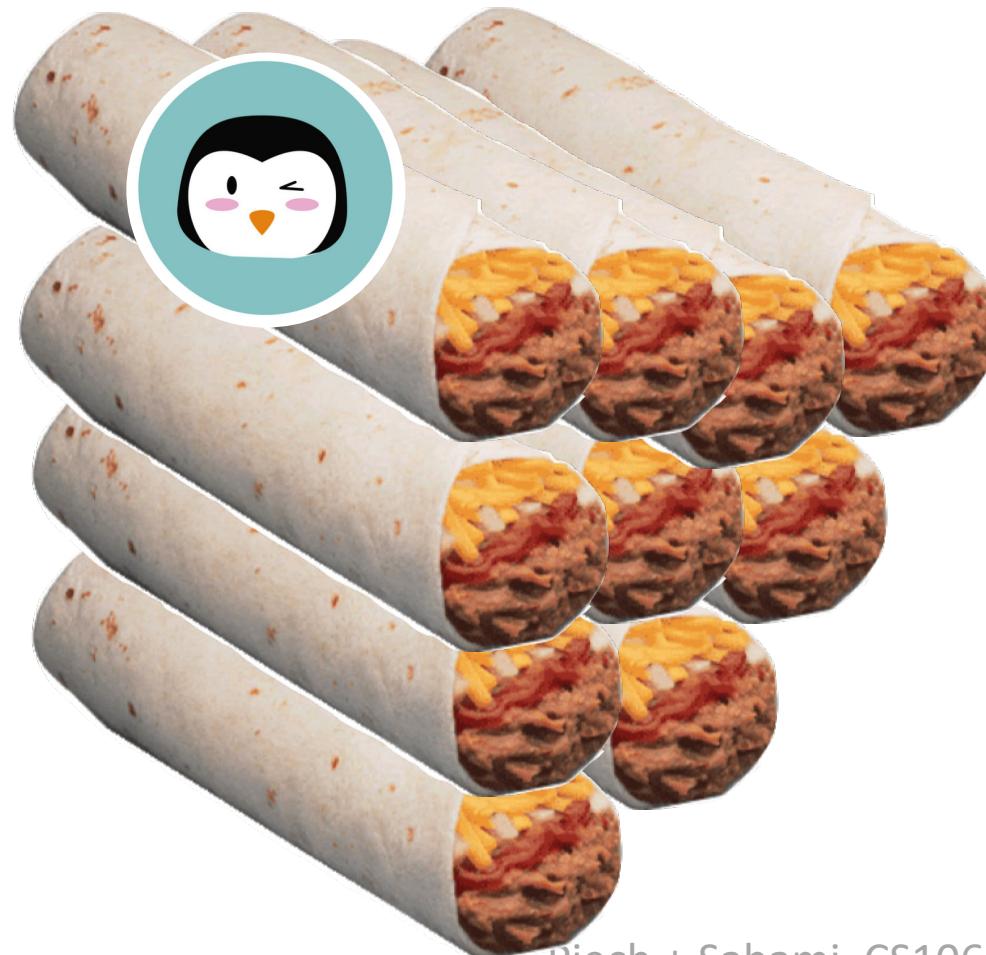
# Distributive Harms

How should things or outcomes be distributed?



# Distributive Harms

Equality of Opportunity: everyone has same opportunity to pursue good thing (burritos)



# Distributive Harms

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



# Distributive Harms

How should things or outcomes be distributed?



**Equality of opportunity:**  
Everyone has same opportunity to pursue good things (burritos)



**Equality of outcome:**  
everyone has same good things (burritos) and same responsibilities



**Equality of welfare:**  
Everyone has well-being/happiness (but only some people need burritos for happiness)

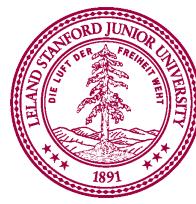
# Representational Harms

Am I represented in this system?

Can I express myself in it?



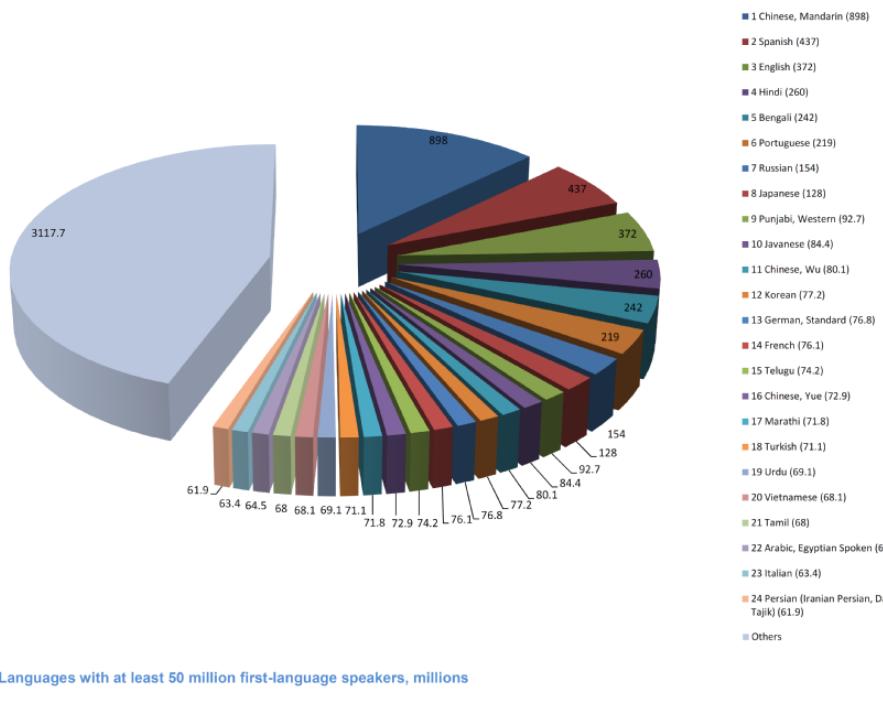
Does this system recognize me, my culture, and my self-expression?



# 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?



# Representation beyond Unicode

What else is needed for linguistic representation besides Unicode?

- **Keyboard support**
  - Texting keyboards!
- **Rendering support**
  - Skin-tone emojis



# Emoji Representation in Unicode

2015: “Family” aka “family\_man\_woman\_boy”

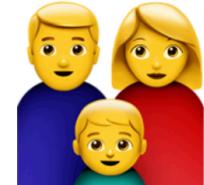
Codepoints

👨 U+1F468  
U+200D  
👩 U+1F469  
U+200D  
👦 U+1F466

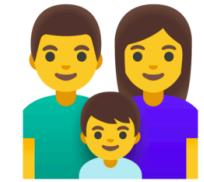
Shortcodes

:family: (Slack)  
:family\_man\_woman\_boy: (Github, Emojipedia)

Apple



Google



Microsoft



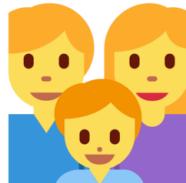
Samsung



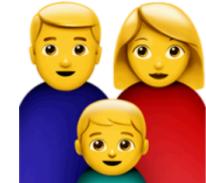
# Emoji Representation in Unicode

2015: “Family” aka “family\_man\_woman\_boy”

Twitter



Apple



Facebook



Google



JoyPixels



Microsoft



OpenMoji



Samsung



# Emoji Representation in Unicode

`"family_man_woman_girl"` `"family_woman_woman_boy"`

Apple



Google



Microsoft



Samsung



Apple



Google



Microsoft



Samsung



# Scientific Representation

Representation of people is different from representation of natural phenomena.

This is a special challenge for social science.

