



**GREG
KONDRAK**

**NOT A HUMAN
LANGUAGE**

1
00:00:04,970 --> 00:00:02,810
you really want to be educated to be

2
00:00:07,610 --> 00:00:04,980
somebody who cannot be replaced by a

3
00:00:09,710 --> 00:00:07,620
computer and I guarantee you that he

4
00:00:11,750 --> 00:00:09,720
will never be able to replace the most

5
00:00:15,490 --> 00:00:11,760
important part of us which is the

6
00:00:15,500 --> 00:00:19,910
are you sure about that

7
00:00:24,109 --> 00:00:21,529
today we talk about the boynage

8
00:00:26,210 --> 00:00:24,119
manuscript the Zodiac CIPHER specialized

9
00:00:28,609 --> 00:00:26,220
non-human languages the Dora Bella

10
00:00:30,710 --> 00:00:28,619
Cipher ciphers in general and how to go

11
00:00:32,330 --> 00:00:30,720
about decrypting them great contract is

12
00:00:33,709 --> 00:00:32,340
a professor of artificial intelligence

13
00:00:35,630 --> 00:00:33,719

investigating natural language

14

00:00:37,370 --> 00:00:35,640

processing in a way that it relates to

15

00:00:38,990 --> 00:00:37,380

language reconstruction he's employed

16

00:00:40,910 --> 00:00:39,000

these machine learning techniques to

17

00:00:43,069 --> 00:00:40,920

attempt what can be considered the most

18

00:00:45,170 --> 00:00:43,079

objective decipherment of the voynage

19

00:00:47,270 --> 00:00:45,180

manuscript which is a considerably rare

20

00:00:48,889 --> 00:00:47,280

Illustrated codex handwritten in an

21

00:00:50,810 --> 00:00:48,899

otherwise unknown writing system it's

22

00:00:52,430 --> 00:00:50,820

evaded any attempt to decode it since

23

00:00:54,049 --> 00:00:52,440

the Italian Renaissance the voynage

24

00:00:56,270 --> 00:00:54,059

manuscript is written on an expensive

25

00:00:57,889 --> 00:00:56,280

Vellum and it's just one of the several

26

00:00:59,750 --> 00:00:57,899

puzzles that Professor contract has

27

00:01:01,610 --> 00:00:59,760

tackled another example being the Dora

28

00:01:03,590 --> 00:01:01,620

Bella Cipher Greg kondrak is also known

29

00:01:05,570 --> 00:01:03,600

for proving chomsky's statement wrong

30

00:01:07,429 --> 00:01:05,580

the statement that English orthography

31

00:01:08,690 --> 00:01:07,439

is close to Optimum my name is Kurt J

32

00:01:10,609 --> 00:01:08,700

mungle I have a background in

33

00:01:11,990 --> 00:01:10,619

mathematical physics this podcast is

34

00:01:14,330 --> 00:01:12,000

called theories of everything is

35

00:01:16,429 --> 00:01:14,340

dedicated to the exploration of theories

36

00:01:18,289 --> 00:01:16,439

of everything from a theoretical physics

37

00:01:19,550 --> 00:01:18,299

perspective but as well as exploring the

38

00:01:21,770 --> 00:01:19,560

world Consciousness has to the

39

00:01:23,690 --> 00:01:21,780

fundamental laws of nature each sponsor

40

00:01:25,609 --> 00:01:23,700

as well as the patrons improves the

41

00:01:27,350 --> 00:01:25,619

quality of the videos drastically it

42

00:01:29,210 --> 00:01:27,360

improves the depth it improves the

43

00:01:30,710 --> 00:01:29,220

frequency and it goes toward paying the

44

00:01:32,149 --> 00:01:30,720

staff for instance someone who's editing

45

00:01:33,950 --> 00:01:32,159

this full time right now and then we

46

00:01:35,149 --> 00:01:33,960

have an operations manager in that vein

47

00:01:37,010 --> 00:01:35,159

I want to thank today's sponsor

48

00:01:38,690 --> 00:01:37,020

brilliant if you're familiar with tow

49

00:01:40,490 --> 00:01:38,700

you're familiar with Brilliance but for

50

00:01:42,530 --> 00:01:40,500

those who don't know Brilliance is a

51
00:01:44,690 --> 00:01:42,540
place where you go to learn math science

52
00:01:46,670 --> 00:01:44,700
and engineering through these bite-sized

53
00:01:48,230 --> 00:01:46,680
Interactive Learning experiences for

54
00:01:49,850 --> 00:01:48,240
example and I keep saying this I would

55
00:01:52,969 --> 00:01:49,860
like to do a podcast on information

56
00:01:55,249 --> 00:01:52,979
Theory particularly Chiara marleto which

57
00:01:56,450 --> 00:01:55,259
is David Deutsch's student has a theory

58
00:01:58,550 --> 00:01:56,460
of everything that she puts forward

59
00:02:00,710 --> 00:01:58,560
called Constructor Theory which is

60
00:02:02,450 --> 00:02:00,720
heavily contingent on information Theory

61
00:02:04,490 --> 00:02:02,460
so I took their course on random

62
00:02:06,590 --> 00:02:04,500
variable distributions and knowledge and

63
00:02:08,570 --> 00:02:06,600

uncertainty in order to learn a bit more

64

00:02:10,370 --> 00:02:08,580

about entropy now there's this formula

65

00:02:12,350 --> 00:02:10,380

for entropy essentially hammered into

66

00:02:14,270 --> 00:02:12,360

you as an undergraduate which seems to

67

00:02:16,010 --> 00:02:14,280

have fallen from the sky however when

68

00:02:17,510 --> 00:02:16,020

you take Brilliance course it was the

69

00:02:19,790 --> 00:02:17,520

first time that I could see that it's an

70

00:02:22,610 --> 00:02:19,800

extremely clear and intuitive form

71

00:02:24,350 --> 00:02:22,620

formula that is to say that it would be

72

00:02:27,350 --> 00:02:24,360

unnatural to Define it in any other

73

00:02:29,510 --> 00:02:27,360

manner visit brilliant.org toe that is

74

00:02:31,309 --> 00:02:29,520

t-o-e to get 20 off the annual

75

00:02:33,890 --> 00:02:31,319

subscription and I recommend that you

76

00:02:35,630 --> 00:02:33,900

don't stop before four lessons I think

77

00:02:37,250 --> 00:02:35,640

you'll be greatly surprised that the

78

00:02:39,350 --> 00:02:37,260

ease at which you can now comprehend

79

00:02:41,330 --> 00:02:39,360

subjects you previously had a difficult

80

00:02:42,589 --> 00:02:41,340

time grocking at some point I'll also go

81

00:02:44,449 --> 00:02:42,599

through the courses and give a

82

00:02:45,530 --> 00:02:44,459

recommendation in order if you'd like to

83

00:02:47,750 --> 00:02:45,540

support the theories of everything

84

00:02:49,670 --> 00:02:47,760

podcast and help assist that is the

85

00:02:51,589 --> 00:02:49,680

improved video quality the depth the

86

00:02:53,390 --> 00:02:51,599

frequency of the podcast and paying the

87

00:02:55,150 --> 00:02:53,400

staff then do consider going to

88

00:02:57,470 --> 00:02:55,160

patreon.com

89

00:02:59,089 --> 00:02:57,480

kurtjimungle the link is on screen as

90

00:03:00,650 --> 00:02:59,099

well as in the description there's a

91

00:03:02,690 --> 00:03:00,660

custom amount as well as already

92

00:03:04,309 --> 00:03:02,700

delineated tiers either way your

93

00:03:06,949 --> 00:03:04,319

viewership is thank you enough enjoy

94

00:03:09,050 --> 00:03:06,959

this conversation with Greg kondrak

95

00:03:11,449 --> 00:03:09,060

Professor what is the voynish manuscript

96

00:03:14,270 --> 00:03:11,459

and why is it important

97

00:03:18,530 --> 00:03:14,280

about a transcript is a medieval

98

00:03:21,290 --> 00:03:18,540

manuscript written in some code that is

99

00:03:25,190 --> 00:03:21,300

was actually confirmed to be

100

00:03:26,830 --> 00:03:25,200

actually genuine manuscript from the

101

00:03:31,250 --> 00:03:26,840

15th century

102

00:03:35,149 --> 00:03:31,260

it is it has illustrations it has text

103

00:03:40,130 --> 00:03:35,159

the script or the alphabet is unique for

104

00:03:40,670 --> 00:03:40,140

the manuscript has not been deciphered

105

00:03:42,830 --> 00:03:40,680

um

106

00:03:46,490 --> 00:03:42,840

yet

107

00:03:50,089 --> 00:03:46,500

why hasn't it been deciphered

108

00:03:52,009 --> 00:03:50,099

uh that's uh not we don't know why it

109

00:03:55,130 --> 00:03:52,019

hasn't been deciphered some people say

110

00:03:57,470 --> 00:03:55,140

it's because there's nothing to decipher

111

00:04:01,610 --> 00:03:57,480

and it's it's just some kind of a joke

112

00:04:04,970 --> 00:04:01,620

or other people say that the encoding

113

00:04:07,009 --> 00:04:04,980

system is very complicated

114

00:04:09,410 --> 00:04:07,019

I mean other people have other theories

115

00:04:11,630 --> 00:04:09,420

about it why do people say it's a joke

116

00:04:12,670 --> 00:04:11,640

and what do you make of that

117

00:04:17,330 --> 00:04:12,680

yeah

118

00:04:19,569 --> 00:04:17,340

well I I personally I don't think uh as

119

00:04:22,069 --> 00:04:19,579

I said that this is a testable

120

00:04:24,530 --> 00:04:22,079

scientific hypothesis

121

00:04:26,270 --> 00:04:24,540

you can you can guess that it's a joke

122

00:04:27,610 --> 00:04:26,280

but it's very hard to prove something

123

00:04:30,590 --> 00:04:27,620

like that

124

00:04:33,110 --> 00:04:30,600

obviously it cost a lot of money to

125

00:04:34,610 --> 00:04:33,120

produce this kind of manuscript in the

126

00:04:37,370 --> 00:04:34,620

Middle Ages so

127

00:04:39,710 --> 00:04:37,380

that's uh one reason that I don't think

128

00:04:46,129 --> 00:04:39,720

it was a joke

129

00:04:47,570 --> 00:04:46,139

um well but also it it's um there's some

130

00:04:50,810 --> 00:04:47,580

words showing that there are some

131

00:04:53,150 --> 00:04:50,820

statistical properties that indicates

132

00:04:54,650 --> 00:04:53,160

there is actually a language that is

133

00:04:56,689 --> 00:04:54,660

being in code

134

00:04:58,730 --> 00:04:56,699

I've seen several documentaries on the

135

00:05:00,830 --> 00:04:58,740

Voynich manuscript several so not just

136

00:05:02,870 --> 00:05:00,840

one or two there's a variety of them and

137

00:05:05,030 --> 00:05:02,880

then there's also a whole subreddit so a

138

00:05:07,430 --> 00:05:05,040

whole Reddit group dedicated to solving

139

00:05:09,770 --> 00:05:07,440

this why is this so difficult compared

140

00:05:12,350 --> 00:05:09,780

to other ciphers in the past like what

141

00:05:17,090 --> 00:05:12,360

is it about this

142

00:05:18,590 --> 00:05:17,100

well the difficulty here is that the

143

00:05:21,310 --> 00:05:18,600

first difficult is that we don't know

144

00:05:23,990 --> 00:05:21,320

what language it is

145

00:05:26,330 --> 00:05:24,000

very often we have ciphers we have

146

00:05:29,629 --> 00:05:26,340

messages but we know what the message is

147

00:05:32,689 --> 00:05:29,639

for example the German Enigma machine

148

00:05:37,010 --> 00:05:32,699

was a very hard Cipher but we knew that

149

00:05:39,409 --> 00:05:37,020

it was German being encoded which made

150

00:05:41,210 --> 00:05:39,419

it quite easier

151
00:05:43,730 --> 00:05:41,220
and the second thing is that we don't

152
00:05:45,650 --> 00:05:43,740
know the script or the alphabet

153
00:05:48,230 --> 00:05:45,660
if that alphabet was used for something

154
00:05:50,629 --> 00:05:48,240
else we would know how to speak it how

155
00:05:54,469 --> 00:05:50,639
to pronounce these words

156
00:05:55,909 --> 00:05:54,479
and third problem is that this is just a

157
00:05:57,070 --> 00:05:55,919
unique document there's no other

158
00:06:00,710 --> 00:05:57,080
document

159
00:06:04,730 --> 00:06:00,720
that is it is written in this way

160
00:06:07,070 --> 00:06:04,740
so it's all self-contained and we don't

161
00:06:08,870 --> 00:06:07,080
even know we're not exactly sure where

162
00:06:10,430 --> 00:06:08,880
it was even produced

163
00:06:12,590 --> 00:06:10,440

it's strange that there's no other

164

00:06:13,969 --> 00:06:12,600

document that's like that firstly that

165

00:06:15,170 --> 00:06:13,979

the script is different like you

166

00:06:16,969 --> 00:06:15,180

mentioned that we don't know how to

167

00:06:19,070 --> 00:06:16,979

pronounce the words though even in the

168

00:06:20,629 --> 00:06:19,080

Enigma code it's not as if the words

169

00:06:21,950 --> 00:06:20,639

were meant to be pronounced and then

170

00:06:23,930 --> 00:06:21,960

understood like that you could have

171

00:06:25,070 --> 00:06:23,940

translated it to zeros and ones and it

172

00:06:27,110 --> 00:06:25,080

still would have been difficult it still

173

00:06:28,490 --> 00:06:27,120

would have been the same problem so why

174

00:06:30,770 --> 00:06:28,500

does the fact that we don't understand

175

00:06:32,090 --> 00:06:30,780

how to pronounce the alphabet make a

176

00:06:33,710 --> 00:06:32,100

difference why can't we just say look

177

00:06:36,170 --> 00:06:33,720

this letter appears let's call that

178

00:06:38,629 --> 00:06:36,180

letter 28 let's call that other letter

179

00:06:40,850 --> 00:06:38,639

50. well yeah well so we have letters

180

00:06:42,710 --> 00:06:40,860

but we have words words are made of

181

00:06:45,409 --> 00:06:42,720

letters that's the truth in every

182

00:06:47,510 --> 00:06:45,419

language human language that you have

183

00:06:51,830 --> 00:06:47,520

words that are made of

184

00:06:53,090 --> 00:06:51,840

phonings they're usually every phone has

185

00:06:57,430 --> 00:06:53,100

its own letter

186

00:07:01,129 --> 00:06:57,440

and then there are certain

187

00:07:03,770 --> 00:07:01,139

patterns regularities for example there

188

00:07:05,809 --> 00:07:03,780

are certain words that can be pronounced

189

00:07:08,029 --> 00:07:05,819

and other words that cannot be

190

00:07:09,290 --> 00:07:08,039

pronounced and that varies from language

191

00:07:13,490 --> 00:07:09,300

to language

192

00:07:16,430 --> 00:07:13,500

so so the moment you you know what

193

00:07:18,430 --> 00:07:16,440

language we're speaking we know you

194

00:07:21,409 --> 00:07:18,440

really know the Lexicon of that language

195

00:07:23,930 --> 00:07:21,419

typically a few thousand words

196

00:07:26,150 --> 00:07:23,940

and they have different frequencies

197

00:07:28,430 --> 00:07:26,160

so that makes it a lot easier to

198

00:07:30,909 --> 00:07:28,440

decipher anything even if you just

199

00:07:32,809 --> 00:07:30,919

replace every word with a number

200

00:07:34,490 --> 00:07:32,819

that will still give you some

201
00:07:37,010 --> 00:07:34,500
information about the frequency of the

202
00:07:38,450 --> 00:07:37,020
words as long as you do something to

203
00:07:39,890 --> 00:07:38,460
compare it to

204
00:07:41,930 --> 00:07:39,900
I know there are different kinds of

205
00:07:43,490 --> 00:07:41,940
ciphers and I believe the simplest is

206
00:07:44,809 --> 00:07:43,500
called the substitution Cipher that's

207
00:07:48,110 --> 00:07:44,819
like when you're a kid and you just

208
00:07:51,469 --> 00:07:48,120
replace the letter A with c and so on

209
00:07:53,749 --> 00:07:51,479
what other kinds of ciphers are there

210
00:07:56,150 --> 00:07:53,759
yeah so the principally there are

211
00:07:59,330 --> 00:07:56,160
substitution ciphers which is replacing

212
00:08:02,570 --> 00:07:59,340
a symbols and transposition ciphers

213
00:08:04,309 --> 00:08:02,580

which is mixing them up changing their

214

00:08:08,390 --> 00:08:04,319

sequence

215

00:08:10,010 --> 00:08:08,400

and every Cipher is a combination of

216

00:08:13,309 --> 00:08:10,020

those two methods

217

00:08:14,450 --> 00:08:13,319

now we don't know what is varnish we

218

00:08:18,309 --> 00:08:14,460

don't know if it's just a simple

219

00:08:20,930 --> 00:08:18,319

substitution or is it a substitution

220

00:08:23,990 --> 00:08:20,940

combined with transposition

221

00:08:26,390 --> 00:08:24,000

the paper that we wrote assumed that the

222

00:08:28,969 --> 00:08:26,400

transposition was involved

223

00:08:33,170 --> 00:08:28,979

and we try to come up with a general

224

00:08:34,969 --> 00:08:33,180

method of breaking this kind of ciphers

225

00:08:36,350 --> 00:08:34,979

that combine substitutions in just

226

00:08:38,570 --> 00:08:36,360

positions

227

00:08:40,909 --> 00:08:38,580

I completely glossed over the history

228

00:08:42,829 --> 00:08:40,919

actually I not that I glossed over I had

229

00:08:45,170 --> 00:08:42,839

forgotten to even ask you to give the

230

00:08:47,329 --> 00:08:45,180

audience an indication as to where was

231

00:08:49,850 --> 00:08:47,339

this document found why does it matter

232

00:08:51,949 --> 00:08:49,860

why is it that Scholars even care about

233

00:08:54,230 --> 00:08:51,959

it I'm sure there's plenty about the

234

00:08:56,090 --> 00:08:54,240

past that we don't know about so why is

235

00:08:57,949 --> 00:08:56,100

it that many scholars and not only

236

00:09:00,230 --> 00:08:57,959

Scholars groups of people teams of

237

00:09:01,730 --> 00:09:00,240

people on Reddit are pouring over trying

238

00:09:03,530 --> 00:09:01,740

to figure out what the heck is this

239

00:09:06,350 --> 00:09:03,540

saying where was it from can you give a

240

00:09:10,810 --> 00:09:06,360

bit of the history of it please

241

00:09:13,970 --> 00:09:10,820

sure so sorry was found basically by the

242

00:09:16,910 --> 00:09:13,980

person whose name was Vonage that's what

243

00:09:18,230 --> 00:09:16,920

it's called the Voynich it was a kind of

244

00:09:22,430 --> 00:09:18,240

a collector

245

00:09:26,090 --> 00:09:22,440

uh in the beginning of the 20th century

246

00:09:29,630 --> 00:09:26,100

and since then the the manuscript has

247

00:09:32,630 --> 00:09:29,640

been tracked back to 17th century to the

248

00:09:34,430 --> 00:09:32,640

Court of uh Roman Emperor and that's

249

00:09:36,170 --> 00:09:34,440

where it kind of ends that's where the

250

00:09:41,050 --> 00:09:36,180

Trail Ends

251

00:09:43,670 --> 00:09:41,060

uh but as uh not long ago there was a

252

00:09:46,009 --> 00:09:43,680

chemical analysis done on the manuscript

253

00:09:49,070 --> 00:09:46,019

so we're certain that it was actually

254

00:09:52,250 --> 00:09:49,080

written in the 15th century

255

00:09:55,130 --> 00:09:52,260

so there is no doubt about that anymore

256

00:09:57,290 --> 00:09:55,140

now you also asked the other question

257

00:09:58,370 --> 00:09:57,300

which is why are people so fascinated

258

00:10:01,370 --> 00:09:58,380

with it

259

00:10:03,410 --> 00:10:01,380

I think the main reason is that we are

260

00:10:06,530 --> 00:10:03,420

fascinated by puzzles if we see

261

00:10:08,930 --> 00:10:06,540

something that seems to be a message we

262

00:10:10,790 --> 00:10:08,940

want to know what that message is we

263

00:10:14,329 --> 00:10:10,800

want to decipher it

264

00:10:15,050 --> 00:10:14,339

and Voyage is like a Mount Everest of

265

00:10:18,250 --> 00:10:15,060

all

266

00:10:21,350 --> 00:10:18,260

uh overloading

267

00:10:23,150 --> 00:10:21,360

cryptographic puzzles

268

00:10:26,810 --> 00:10:23,160

it was actually

269

00:10:30,110 --> 00:10:26,820

studied by many years by people that

270

00:10:32,690 --> 00:10:30,120

were working for the U.S government

271

00:10:35,570 --> 00:10:32,700

and that they were professional Breakers

272

00:10:37,850 --> 00:10:35,580

code Breakers that that broke cause

273

00:10:40,550 --> 00:10:37,860

during the second World War

274

00:10:41,990 --> 00:10:40,560

in the case of the Enigma code we at

275

00:10:43,730 --> 00:10:42,000

least we could presume that it has

276

00:10:46,250 --> 00:10:43,740

something to do with government secrets

277

00:10:47,810 --> 00:10:46,260

and War whereas with this do we have any

278

00:10:49,730 --> 00:10:47,820

indication as to what the subject matter

279

00:10:51,769 --> 00:10:49,740

is

280

00:10:54,829 --> 00:10:51,779

well we have the illustrations that's

281

00:11:00,230 --> 00:10:54,839

what makes it really interesting is and

282

00:11:06,769 --> 00:11:04,069

all kinds of strange illustrations that

283

00:11:10,009 --> 00:11:06,779

seem to be related to the text

284

00:11:11,930 --> 00:11:10,019

so so this is not just a text like some

285

00:11:14,329 --> 00:11:11,940

kind of ancient inscription but it is

286

00:11:16,550 --> 00:11:14,339

actually a codex which is like a

287

00:11:19,190 --> 00:11:16,560

compendium of some kind of knowledge

288

00:11:21,710 --> 00:11:19,200

which was quite common in Middle Ages

289

00:11:23,090 --> 00:11:21,720

and is it quite common among a

290

00:11:24,590 --> 00:11:23,100

particular group that speaks a

291

00:11:26,210 --> 00:11:24,600

particular language and thus we could

292

00:11:28,970 --> 00:11:26,220

figure out okay with this amount of

293

00:11:30,970 --> 00:11:28,980

probability it's from this culture or

294

00:11:35,690 --> 00:11:30,980

this group of people

295

00:11:38,210 --> 00:11:35,700

yeah so many of those codex codices are

296

00:11:41,449 --> 00:11:38,220

in Latin which was the language

297

00:11:44,810 --> 00:11:41,459

of literature and Science in Middle Ages

298

00:11:48,350 --> 00:11:44,820

and you can actually find such such

299

00:11:50,569 --> 00:11:48,360

books very similar similar looking for

300

00:11:53,870 --> 00:11:50,579

Middle Ages that are written in Latin

301
00:11:56,930 --> 00:11:53,880
the best guess is about the provenance

302
00:11:58,310 --> 00:11:56,940
of the manuscripts uh point at Northern

303
00:12:01,850 --> 00:11:58,320
Italy

304
00:12:04,069 --> 00:12:01,860
were of course Latin was the language of

305
00:12:06,170 --> 00:12:04,079
literature as well at that time

306
00:12:08,449 --> 00:12:06,180
is it controversial that it's in the

307
00:12:10,970 --> 00:12:08,459
15th century and in Northern Italy

308
00:12:14,030 --> 00:12:10,980
almost everything that you can say about

309
00:12:16,610 --> 00:12:14,040
foreign so

310
00:12:18,110 --> 00:12:16,620
this is what I consider a reasonable

311
00:12:19,850 --> 00:12:18,120
guest but

312
00:12:22,430 --> 00:12:19,860
and pretty much everybody has a

313
00:12:24,889 --> 00:12:22,440

different opinion where this uh this

314

00:12:26,509 --> 00:12:24,899

manuscript comes from and what language

315

00:12:29,389 --> 00:12:26,519

it represents

316

00:12:32,990 --> 00:12:29,399

I I don't have any it is not my research

317

00:12:36,230 --> 00:12:33,000

that points to the Northern Italy

318

00:12:37,850 --> 00:12:36,240

I only looked at this manuscript from

319

00:12:39,350 --> 00:12:37,860

the point of view of computation

320

00:12:41,269 --> 00:12:39,360

decipherment

321

00:12:43,310 --> 00:12:41,279

speaking of your research now would be a

322

00:12:44,870 --> 00:12:43,320

great time to tell the audience what is

323

00:12:46,069 --> 00:12:44,880

it that you study and then how the heck

324

00:12:48,170 --> 00:12:46,079

did you become interested in the

325

00:12:51,230 --> 00:12:48,180

voendage manuscript other than a general

326

00:12:54,410 --> 00:12:51,240

curiosity for solving puzzles

327

00:12:57,050 --> 00:12:54,420

yeah so I I'm a computational linguist

328

00:12:59,509 --> 00:12:57,060

I'd say so I I work at the computer

329

00:13:01,730 --> 00:12:59,519

science department at the University of

330

00:13:04,610 --> 00:13:01,740

Alberta in Canada

331

00:13:07,310 --> 00:13:04,620

and I work on language in general making

332

00:13:10,069 --> 00:13:07,320

computers understand human language

333

00:13:12,350 --> 00:13:10,079

making a rewriting programs that can

334

00:13:15,110 --> 00:13:12,360

process human language and do the work

335

00:13:17,269 --> 00:13:15,120

for us because there's so much text

336

00:13:19,910 --> 00:13:17,279

that is available that nobody can

337

00:13:23,090 --> 00:13:19,920

actually read that all of that all of it

338

00:13:25,670 --> 00:13:23,100

about the the decipherment

339

00:13:27,650 --> 00:13:25,680

the person that made this really

340

00:13:29,930 --> 00:13:27,660

interesting for me was Professor Kevin

341

00:13:31,370 --> 00:13:29,940

knight from University of Southern

342

00:13:35,449 --> 00:13:31,380

California

343

00:13:37,970 --> 00:13:35,459

and he worked on various interesting

344

00:13:39,650 --> 00:13:37,980

projects and I saw his presentation of

345

00:13:42,710 --> 00:13:39,660

wineish manuscript

346

00:13:45,949 --> 00:13:42,720

about 10 years ago I I say

347

00:13:46,730 --> 00:13:45,959

and it it was related to what I was

348

00:13:50,269 --> 00:13:46,740

doing

349

00:13:52,490 --> 00:13:50,279

what Kevin Dr Knight said basically is

350

00:13:54,350 --> 00:13:52,500

that everything we do with language is a

351

00:13:56,329 --> 00:13:54,360

kind of decipherment

352

00:13:59,329 --> 00:13:56,339

because language

353

00:14:01,370 --> 00:13:59,339

we is typically written that's what we

354

00:14:03,889 --> 00:14:01,380

work with a written language even if

355

00:14:06,949 --> 00:14:03,899

it's spoken we work we work with some

356

00:14:08,930 --> 00:14:06,959

form of it that is a transcription

357

00:14:11,690 --> 00:14:08,940

and whenever you have a sequence of

358

00:14:15,290 --> 00:14:11,700

symbols it becomes a decipherment

359

00:14:17,269 --> 00:14:15,300

problem basically deciphering first what

360

00:14:19,430 --> 00:14:17,279

are the phonemes behind those symbols

361

00:14:20,870 --> 00:14:19,440

and secondly is the meaning behind the

362

00:14:22,670 --> 00:14:20,880

symbols

363

00:14:25,069 --> 00:14:22,680

with respect to the images in the

364

00:14:26,990 --> 00:14:25,079

Voynich which will overlay on screen is

365

00:14:29,090 --> 00:14:27,000

it strange to depict what they depict

366

00:14:30,410 --> 00:14:29,100

the pictures yeah yeah like what are the

367

00:14:32,470 --> 00:14:30,420

pictures of and is there something

368

00:14:36,230 --> 00:14:32,480

unique about them

369

00:14:38,810 --> 00:14:36,240

yes so so the generally if I tell you

370

00:14:41,030 --> 00:14:38,820

that it depicts plans for example that's

371

00:14:43,850 --> 00:14:41,040

not strange because that's what the

372

00:14:46,790 --> 00:14:43,860

medieval codices do but if you look

373

00:14:49,370 --> 00:14:46,800

closely at those plants if you're an

374

00:14:51,650 --> 00:14:49,380

expert in Plants I am not but

375

00:14:53,269 --> 00:14:51,660

experts on Plants look at them and say

376

00:14:55,550 --> 00:14:53,279

well these don't really look like real

377

00:14:56,750 --> 00:14:55,560

plants these are look like made up

378

00:14:59,449 --> 00:14:56,760

plants

379

00:15:04,490 --> 00:14:59,459

and then there are pictures of people

380

00:15:06,230 --> 00:15:04,500

that some kind of many naked bodies

381

00:15:09,170 --> 00:15:06,240

taking baths

382

00:15:11,990 --> 00:15:09,180

in some kind of green water

383

00:15:14,329 --> 00:15:12,000

you know in general pictures of people

384

00:15:16,310 --> 00:15:14,339

are not strange but those particular

385

00:15:19,430 --> 00:15:16,320

pictures are really strange

386

00:15:21,290 --> 00:15:19,440

and they are unlike anything else that

387

00:15:23,150 --> 00:15:21,300

we know from middle age

388

00:15:25,370 --> 00:15:23,160

they're strange because they're naked or

389

00:15:27,350 --> 00:15:25,380

they're strange because they're depicted

390

00:15:29,990 --> 00:15:27,360

alongside plants like what is it

391

00:15:32,090 --> 00:15:30,000

specifically that's unique

392

00:15:34,490 --> 00:15:32,100

The Exchange because it's not clear what

393

00:15:37,370 --> 00:15:34,500

they are depicting are they depicting

394

00:15:39,650 --> 00:15:37,380

people taking bath that I don't think

395

00:15:43,610 --> 00:15:39,660

was very common in Middle Ages

396

00:15:46,550 --> 00:15:43,620

or why are why are those figures for

397

00:15:49,910 --> 00:15:46,560

example all women and why they are naked

398

00:15:53,930 --> 00:15:49,920

right in 15th century that that was not

399

00:15:55,850 --> 00:15:53,940

a normal thing to put in a book

400

00:15:59,509 --> 00:15:55,860

um there are also other things like

401
00:16:01,069 --> 00:15:59,519
zodiac signs or or pictures of plan of

402
00:16:04,389 --> 00:16:01,079
of planets

403
00:16:07,250 --> 00:16:04,399
that you would expect to be

404
00:16:10,189 --> 00:16:07,260
to be quite normal because we even know

405
00:16:14,090 --> 00:16:10,199
those the what those Zodiacs are

406
00:16:16,009 --> 00:16:14,100
but it's it's difficult to connect the

407
00:16:18,590 --> 00:16:16,019
words that describes those pictures to

408
00:16:20,629 --> 00:16:18,600
the actual pictures

409
00:16:22,790 --> 00:16:20,639
in one of the documentaries that I was

410
00:16:25,389 --> 00:16:22,800
watching about this they said that a

411
00:16:29,090 --> 00:16:25,399
remarkable element is that there are

412
00:16:31,310 --> 00:16:29,100
extremely few errors in a writing of

413
00:16:32,509 --> 00:16:31,320

this size they would expect that there

414

00:16:34,250 --> 00:16:32,519

are some errors and then maybe you

415

00:16:35,870 --> 00:16:34,260

smudge it out or however they correct

416

00:16:37,490 --> 00:16:35,880

errors and there's some way of detecting

417

00:16:39,769 --> 00:16:37,500

the frequency of the errors in a

418

00:16:42,230 --> 00:16:39,779

document and most documents have let's

419

00:16:44,629 --> 00:16:42,240

say error percentage two two percents

420

00:16:48,170 --> 00:16:44,639

they make an error every two out of 100

421

00:16:50,210 --> 00:16:48,180

words whereas this it's much lower like

422

00:16:51,470 --> 00:16:50,220

half of that or even half of half of

423

00:16:53,689 --> 00:16:51,480

that first thing I want to know is that

424

00:16:55,490 --> 00:16:53,699

even true and then secondly like why why

425

00:16:57,470 --> 00:16:55,500

do you think that's the case what could

426
00:16:59,749 --> 00:16:57,480
that mean

427
00:17:01,850 --> 00:16:59,759
yeah so first of all I did not work with

428
00:17:04,250 --> 00:17:01,860
the actual text I worked with the

429
00:17:07,429 --> 00:17:04,260
transcription that somebody made

430
00:17:09,829 --> 00:17:07,439
but it is from what I know it is true

431
00:17:12,289 --> 00:17:09,839
that there is very few Corrections

432
00:17:15,409 --> 00:17:12,299
my personal opinion that it may indicate

433
00:17:17,449 --> 00:17:15,419
that whoever was writing this copying it

434
00:17:19,970 --> 00:17:17,459
did not understand what they were

435
00:17:21,710 --> 00:17:19,980
writing you usually make Corrections if

436
00:17:24,110 --> 00:17:21,720
you write something and you see oh it's

437
00:17:25,429 --> 00:17:24,120
that's not what it should be right but

438
00:17:28,429 --> 00:17:25,439

if you write something in a language

439

00:17:30,110 --> 00:17:28,439

that you're totally unfamiliar with you

440

00:17:32,090 --> 00:17:30,120

won't be able to notice that that's

441

00:17:34,370 --> 00:17:32,100

interesting in that case that would

442

00:17:38,510 --> 00:17:34,380

imply that there's another copy no that

443

00:17:42,049 --> 00:17:38,520

means that I I there were people that

444

00:17:46,490 --> 00:17:42,059

maybe were copying the text

445

00:17:48,470 --> 00:17:46,500

into the manuscript because that was a

446

00:17:50,690 --> 00:17:48,480

you would expect there was some draft

447

00:17:53,450 --> 00:17:50,700

that we're copying from because the

448

00:17:56,029 --> 00:17:53,460

manuscript itself was very expensive uh

449

00:17:58,070 --> 00:17:56,039

to write down and we also know that

450

00:17:59,990 --> 00:17:58,080

there are different kind of hands so it

451
00:18:01,909 --> 00:18:00,000
looks like they were more than a one

452
00:18:04,370 --> 00:18:01,919
person doing the copying that's

453
00:18:06,350 --> 00:18:04,380
interesting then that means that it's

454
00:18:08,630 --> 00:18:06,360
extra difficult to decipher it because

455
00:18:10,310 --> 00:18:08,640
there are errors we don't know this but

456
00:18:11,870 --> 00:18:10,320
that would mean that there's more than

457
00:18:13,909 --> 00:18:11,880
the average amount of Errors what's the

458
00:18:15,350 --> 00:18:13,919
difference between inciphering so

459
00:18:17,810 --> 00:18:15,360
creating a cipher out of something and

460
00:18:19,669 --> 00:18:17,820
then encryption

461
00:18:21,830 --> 00:18:19,679
yeah I'm not sure if there is a

462
00:18:23,330 --> 00:18:21,840
difference you know if you really try to

463
00:18:24,830 --> 00:18:23,340

find the difference you could say that

464

00:18:27,590 --> 00:18:24,840

encryption is

465

00:18:29,450 --> 00:18:27,600

like you do encryption of a text but you

466

00:18:32,990 --> 00:18:29,460

don't really know how it is done you

467

00:18:35,690 --> 00:18:33,000

know you apply some encryption program

468

00:18:37,789 --> 00:18:35,700

whereas in ciphering implies that you go

469

00:18:40,130 --> 00:18:37,799

like letter by letter look up the key

470

00:18:41,570 --> 00:18:40,140

and then Cipher each each letter

471

00:18:44,210 --> 00:18:41,580

separately

472

00:18:45,950 --> 00:18:44,220

but in general I think it's it's pretty

473

00:18:49,130 --> 00:18:45,960

much the same thing

474

00:18:51,830 --> 00:18:49,140

what were the different methods used to

475

00:18:53,750 --> 00:18:51,840

decipher the Voynich manuscript not just

476
00:18:55,970 --> 00:18:53,760
by you but by others what techniques do

477
00:18:58,970 --> 00:18:55,980
they employ

478
00:19:02,150 --> 00:18:58,980
well that's a kind of difficult to say

479
00:19:05,210 --> 00:19:02,160
because none of those methods actually

480
00:19:08,750 --> 00:19:05,220
worked and then deciphering has not been

481
00:19:10,990 --> 00:19:08,760
achieved in spite of many claims

482
00:19:14,330 --> 00:19:11,000
to the country so

483
00:19:16,490 --> 00:19:14,340
the there isn't really any algorithm

484
00:19:20,110 --> 00:19:16,500
involved it's it's mostly based on

485
00:19:24,529 --> 00:19:20,120
people's intuition and and theories

486
00:19:27,650 --> 00:19:24,539
but but even then uh if even if there

487
00:19:31,730 --> 00:19:27,660
was some method to it it has not

488
00:19:33,590 --> 00:19:31,740

produced a readable transcription

489

00:19:34,789 --> 00:19:33,600

what are some of your attempts to

490

00:19:37,730 --> 00:19:34,799

decipher can you go through the

491

00:19:41,330 --> 00:19:37,740

successes and failures

492

00:19:43,070 --> 00:19:41,340

yes so so uh in our project our

493

00:19:46,310 --> 00:19:43,080

assumption was that the first thing to

494

00:19:49,430 --> 00:19:46,320

basically to start with is to find out

495

00:19:50,750 --> 00:19:49,440

what language this is written in if we

496

00:19:52,310 --> 00:19:50,760

don't know what the language it is

497

00:19:53,690 --> 00:19:52,320

written and then there's no way to

498

00:19:59,289 --> 00:19:53,700

decipher it

499

00:20:02,150 --> 00:19:59,299

so we devised some methods of detecting

500

00:20:04,970 --> 00:20:02,160

identifying the language of the cipher

501
00:20:08,510 --> 00:20:04,980
even without deciphering it

502
00:20:10,909 --> 00:20:08,520
and we use the sample a large sample of

503
00:20:13,909 --> 00:20:10,919
about 400 languages

504
00:20:16,070 --> 00:20:13,919
and out of those far-handed languages we

505
00:20:17,210 --> 00:20:16,080
assign like a number as score to each

506
00:20:20,150 --> 00:20:17,220
language

507
00:20:21,409 --> 00:20:20,160
in terms of the probability that this

508
00:20:23,210 --> 00:20:21,419
language is the language of the

509
00:20:26,870 --> 00:20:23,220
manuscript

510
00:20:29,150 --> 00:20:26,880
and what was interesting that we found

511
00:20:31,490 --> 00:20:29,160
was that the language that was the the

512
00:20:33,830 --> 00:20:31,500
highest scoring language or out of those

513
00:20:35,990 --> 00:20:33,840

400 was Hebrew

514

00:20:38,150 --> 00:20:36,000

how much higher was it than the second

515

00:20:41,450 --> 00:20:38,160

and third place

516

00:20:43,130 --> 00:20:41,460

was a clear difference I would say a

517

00:20:44,870 --> 00:20:43,140

significant difference between the

518

00:20:47,930 --> 00:20:44,880

second one in the list

519

00:20:50,690 --> 00:20:47,940

so that was quite striking

520

00:20:52,850 --> 00:20:50,700

and were you able to get any other

521

00:20:55,850 --> 00:20:52,860

historical documents that are written in

522

00:20:57,830 --> 00:20:55,860

Hebrew that have a similar art style and

523

00:20:59,450 --> 00:20:57,840

are of similar length just to see well

524

00:21:01,430 --> 00:20:59,460

is this common is this a common practice

525

00:21:05,210 --> 00:21:01,440

to the people who write in Hebrew or is

526

00:21:08,630 --> 00:21:05,220

this aberrant is this extremely unique

527

00:21:10,789 --> 00:21:08,640

so the Hebrew manuscripts exist and they

528

00:21:11,750 --> 00:21:10,799

were written throughout Middle Ages in

529

00:21:15,049 --> 00:21:11,760

Hebrew

530

00:21:18,710 --> 00:21:15,059

by by the Jewish Scholars

531

00:21:21,470 --> 00:21:18,720

and I'm not the only person that that uh

532

00:21:24,789 --> 00:21:21,480

that hypothesized that this this was

533

00:21:27,169 --> 00:21:24,799

actually coming from from the Jewish

534

00:21:30,289 --> 00:21:27,179

scholar community

535

00:21:32,049 --> 00:21:30,299

uh now nobody used this kind of

536

00:21:34,990 --> 00:21:32,059

particular script

537

00:21:39,590 --> 00:21:35,000

but this script

538

00:21:42,710 --> 00:21:39,600

does have some some similarities to

539

00:21:47,570 --> 00:21:42,720

Hebrew script for example

540

00:21:49,909 --> 00:21:47,580

I actually don't speak here uh or or I

541

00:21:51,590 --> 00:21:49,919

don't know much about it but I know that

542

00:21:54,289 --> 00:21:51,600

the Hebrew script does not include

543

00:21:55,970 --> 00:21:54,299

letters sorry include vowels which makes

544

00:21:58,610 --> 00:21:55,980

the words shorter

545

00:22:01,730 --> 00:21:58,620

and this is what we observe in vanished

546

00:22:04,970 --> 00:22:01,740

is that the words are quite short

547

00:22:08,149 --> 00:22:04,980

and then the number uh the number of uh

548

00:22:10,970 --> 00:22:08,159

different symbols suggest that it is

549

00:22:13,130 --> 00:22:10,980

something like a substitution Cipher

550

00:22:14,930 --> 00:22:13,140

because the number of symbol is similar

551
00:22:16,010 --> 00:22:14,940
to the number of phonemes in a typical

552
00:22:18,169 --> 00:22:16,020
language

553
00:22:19,490 --> 00:22:18,179
okay so now that you have at least

554
00:22:22,310 --> 00:22:19,500
potentially identified the language

555
00:22:25,789 --> 00:22:22,320
what's the next step

556
00:22:28,250 --> 00:22:25,799
yeah a very good point so the next point

557
00:22:30,649 --> 00:22:28,260
the next step is obviously try to match

558
00:22:32,690 --> 00:22:30,659
every symbol to a different letter of

559
00:22:34,490 --> 00:22:32,700
the say Hebrew

560
00:22:38,330 --> 00:22:34,500
alphabet

561
00:22:41,390 --> 00:22:38,340
and that usually is easy the breaking

562
00:22:43,789 --> 00:22:41,400
simple substitution Cipher's disease but

563
00:22:47,350 --> 00:22:43,799

it doesn't work in this case it does not

564

00:22:50,570 --> 00:22:47,360

produce any sensible decipherment

565

00:22:53,270 --> 00:22:50,580

so we came up with this hypothesis that

566

00:22:56,570 --> 00:22:53,280

the letters within words are actually

567

00:22:58,730 --> 00:22:56,580

transposed to make it more difficult

568

00:23:02,450 --> 00:22:58,740

to decipher

569

00:23:04,669 --> 00:23:02,460

and when you kind of move the letters

570

00:23:06,770 --> 00:23:04,679

around it becomes very difficult to

571

00:23:08,510 --> 00:23:06,780

decipher it so we came up with a method

572

00:23:10,070 --> 00:23:08,520

that could handle this kind of

573

00:23:12,950 --> 00:23:10,080

transposition

574

00:23:16,570 --> 00:23:12,960

within words and we tested it on other

575

00:23:20,029 --> 00:23:16,580

languages and it worked very well

576

00:23:22,430 --> 00:23:20,039

however when we apply this to uh the

577

00:23:25,190 --> 00:23:22,440

varnish manuscript it still does not

578

00:23:26,750 --> 00:23:25,200

produce any kind of a readable

579

00:23:28,130 --> 00:23:26,760

decipherment

580

00:23:30,549 --> 00:23:28,140

when you're testing it with other

581

00:23:33,169 --> 00:23:30,559

languages are you testing it with

582

00:23:35,330 --> 00:23:33,179

encipherments that you contrive or are

583

00:23:38,630 --> 00:23:35,340

you testing it with ciphers that already

584

00:23:41,870 --> 00:23:38,640

exist from those other languages

585

00:23:44,330 --> 00:23:41,880

no we tested it on a mass scale with uh

586

00:23:46,130 --> 00:23:44,340

with synthetic Cipher so computer

587

00:23:48,049 --> 00:23:46,140

generated ciphers

588

00:23:50,149 --> 00:23:48,059

but these are generated from the actual

589

00:23:52,250 --> 00:23:50,159

text in those languages

590

00:23:53,990 --> 00:23:52,260

you mentioned that there's two kinds of

591

00:23:55,730 --> 00:23:54,000

ciphers at least so far so there's

592

00:23:58,250 --> 00:23:55,740

substitution and then transposed or

593

00:24:00,169 --> 00:23:58,260

transposition what else is there so Pig

594

00:24:03,830 --> 00:24:00,179

Latin where you just you add some words

595

00:24:09,950 --> 00:24:07,270

no I think Pig Landing is like a game

596

00:24:11,450 --> 00:24:09,960

what I'm saying is that it seems like

597

00:24:12,950 --> 00:24:11,460

there are other methods that exist even

598

00:24:15,409 --> 00:24:12,960

if they're silly so there's

599

00:24:17,750 --> 00:24:15,419

transposition there's substitution is

600

00:24:19,730 --> 00:24:17,760

that primarily it or is there a seldom

601
00:24:22,549 --> 00:24:19,740
third

602
00:24:24,770 --> 00:24:22,559
well I would say even big Latin you can

603
00:24:28,310 --> 00:24:24,780
express it probably as some kind of

604
00:24:31,730 --> 00:24:28,320
substitution uh or and transposition

605
00:24:35,510 --> 00:24:31,740
right so so these serious ciphers like

606
00:24:37,730 --> 00:24:35,520
in Enigma is is basically again a

607
00:24:41,090 --> 00:24:37,740
combination of substitution and intense

608
00:24:42,649 --> 00:24:41,100
position uh actually Enigma is just pure

609
00:24:45,049 --> 00:24:42,659
substitution really there's no

610
00:24:46,909 --> 00:24:45,059
transposition there except that of

611
00:24:48,470 --> 00:24:46,919
course the spaces between words are

612
00:24:50,690 --> 00:24:48,480
removed

613
00:24:52,669 --> 00:24:50,700

well that's disappointing you're like

614

00:24:54,289 --> 00:24:52,679

okay great I've made a Headway I found

615

00:24:56,270 --> 00:24:54,299

out that it's Hebrew at least you're

616

00:24:58,909 --> 00:24:56,280

somewhat confident it's Hebrew then you

617

00:25:01,130 --> 00:24:58,919

say well okay let me devise some way of

618

00:25:03,350 --> 00:25:01,140

deciphering any substitution plus

619

00:25:05,510 --> 00:25:03,360

transposition combination it works on

620

00:25:07,430 --> 00:25:05,520

other ciphers great let me apply it here

621

00:25:09,350 --> 00:25:07,440

doesn't work so now what are you

622

00:25:11,750 --> 00:25:09,360

thinking and what's next

623

00:25:14,090 --> 00:25:11,760

well first of all I'm not confident that

624

00:25:17,149 --> 00:25:14,100

it is actually Hebrew all I can say is

625

00:25:20,750 --> 00:25:17,159

that out of those 400 languages that we

626
00:25:22,310 --> 00:25:20,760
had samples of this is the one that got

627
00:25:24,950 --> 00:25:22,320
the highest score

628
00:25:28,010 --> 00:25:24,960
so if I had to pick one of those 400

629
00:25:30,769 --> 00:25:28,020
then I would pick here but the language

630
00:25:33,710 --> 00:25:30,779
may actually not not be in that 400

631
00:25:35,450 --> 00:25:33,720
sample it may be a there's thousands of

632
00:25:37,850 --> 00:25:35,460
languages in the world

633
00:25:40,149 --> 00:25:37,860
and in addition it may not actually be

634
00:25:42,409 --> 00:25:40,159
actually any human language some people

635
00:25:45,590 --> 00:25:42,419
hypothesize that it's a made-up language

636
00:25:50,390 --> 00:25:45,600
like Esperanto can this thing

637
00:25:53,149 --> 00:25:50,400
so of course we were excited to see some

638
00:25:54,970 --> 00:25:53,159

kind of clear preference for one of the

639

00:25:56,769 --> 00:25:54,980

languages

640

00:26:01,010 --> 00:25:56,779

but

641

00:26:02,049 --> 00:26:01,020

and and we applied a kind of a

642

00:26:04,909 --> 00:26:02,059

scientific

643

00:26:06,649 --> 00:26:04,919

methodology to it so we reported those

644

00:26:09,409 --> 00:26:06,659

results and they are replicable if

645

00:26:11,450 --> 00:26:09,419

somebody else applies this to that

646

00:26:13,370 --> 00:26:11,460

sample they will find exactly the same

647

00:26:16,850 --> 00:26:13,380

thing but that doesn't mean that this

648

00:26:23,870 --> 00:26:20,029

so uh yeah

649

00:26:26,269 --> 00:26:23,880

if if I was really convinced that this

650

00:26:27,649 --> 00:26:26,279

was Hebrew I think the next thing I

651
00:26:29,990 --> 00:26:27,659
would have to do is to actually learn

652
00:26:32,450 --> 00:26:30,000
Hebrew because

653
00:26:36,669 --> 00:26:32,460
that would be the only way to decipher

654
00:26:40,130 --> 00:26:36,679
that complicated uh manuscript followers

655
00:26:43,190 --> 00:26:40,140
but of course uh I have a lot of other

656
00:26:45,409 --> 00:26:43,200
projects to do so I'm not going to study

657
00:26:48,230 --> 00:26:45,419
Hebrew for that purpose but there are

658
00:26:50,090 --> 00:26:48,240
many people that know Hebrew and uh

659
00:26:52,490 --> 00:26:50,100
and I'm sure if if this was really

660
00:26:53,390 --> 00:26:52,500
Hebrew they would be able to to decipher

661
00:26:56,230 --> 00:26:53,400
it

662
00:26:58,730 --> 00:26:56,240
and themselves so if someone watching

663
00:27:00,710 --> 00:26:58,740

speaks Hebrew and is a computer

664

00:27:03,230 --> 00:27:00,720

scientist and they also want to help

665

00:27:07,570 --> 00:27:03,240

what should they do contact you or is

666

00:27:11,990 --> 00:27:10,130

because I've been contacted by so many

667

00:27:14,390 --> 00:27:12,000

people I'm just gonna put on screen

668

00:27:17,409 --> 00:27:14,400

please contact crondrak here's his email

669

00:27:20,870 --> 00:27:17,419

address and his private phone number

670

00:27:24,049 --> 00:27:20,880

yeah I uh some of those people

671

00:27:26,330 --> 00:27:24,059

uh were actually experts in Hebrew and

672

00:27:28,909 --> 00:27:26,340

in computers and in ciphers

673

00:27:30,950 --> 00:27:28,919

and even they they could not make any

674

00:27:32,570 --> 00:27:30,960

progression so then why do you think

675

00:27:34,250 --> 00:27:32,580

that if you were to learn Hebrew would

676

00:27:38,269 --> 00:27:34,260

help

677

00:27:40,250 --> 00:27:38,279

no I said that if I really was 100 sure

678

00:27:43,370 --> 00:27:40,260

that it's Hebrew then that would

679

00:27:45,769 --> 00:27:43,380

definitely help to know Hebrew right my

680

00:27:48,289 --> 00:27:45,779

work is from a point of view of a

681

00:27:52,310 --> 00:27:48,299

computer scientist not from a point of

682

00:27:54,890 --> 00:27:52,320

view of a linguist or or a cryptographer

683

00:27:57,470 --> 00:27:54,900

so it's not as simple as saying identify

684

00:28:00,549 --> 00:27:57,480

the language then suggest the different

685

00:28:03,950 --> 00:28:00,559

rules so that is the substitution slash

686

00:28:07,549 --> 00:28:03,960

transposition combination and once you

687

00:28:10,610 --> 00:28:07,559

have that feed a text any text that is

688

00:28:12,529 --> 00:28:10,620

in that language and is some unknown

689

00:28:15,230 --> 00:28:12,539

substitution slash transposition

690

00:28:18,710 --> 00:28:15,240

combination and outputs it tells you yes

691

00:28:22,250 --> 00:28:18,720

or no it's not as simple as that

692

00:28:24,230 --> 00:28:22,260

so you know as I said the the main value

693

00:28:28,010 --> 00:28:24,240

of vines manuscript is that it forces

694

00:28:30,590 --> 00:28:28,020

you to come up with new methods that

695

00:28:31,730 --> 00:28:30,600

later may turn out to be useful for

696

00:28:34,669 --> 00:28:31,740

other things

697

00:28:37,970 --> 00:28:34,679

what we came up with is is a methodology

698

00:28:39,350 --> 00:28:37,980

for doing this and it we proved it in

699

00:28:40,970 --> 00:28:39,360

our paper

700

00:28:43,750 --> 00:28:40,980

that it works

701

00:28:47,390 --> 00:28:43,760

maybe at kind of 95

702

00:28:49,310 --> 00:28:47,400

accuracy if you take a language whatever

703

00:28:53,830 --> 00:28:49,320

language speak any language

704

00:28:58,730 --> 00:28:56,690

substitute letters for other symbols

705

00:29:01,430 --> 00:28:58,740

scramble them

706

00:29:03,890 --> 00:29:01,440

give it to our program it will decipher

707

00:29:06,649 --> 00:29:03,900

it with 95 percent like

708

00:29:08,630 --> 00:29:06,659

so that so that is proven and that is a

709

00:29:12,110 --> 00:29:08,640

replicable thing that

710

00:29:13,370 --> 00:29:12,120

was published in the paper

711

00:29:19,010 --> 00:29:13,380

but

712

00:29:20,210 --> 00:29:19,020

that it is actually a actual human

713

00:29:26,090 --> 00:29:20,220

language

714

00:29:27,950 --> 00:29:26,100

is being used for that purpose

715

00:29:31,970 --> 00:29:27,960

the fact that it doesn't work with

716

00:29:34,190 --> 00:29:31,980

Voynich suggests that Vonage is not

717

00:29:35,269 --> 00:29:34,200

written in Hebrew or any of those 400

718

00:29:37,430 --> 00:29:35,279

languages

719

00:29:39,230 --> 00:29:37,440

so you ended up testing it on all 400

720

00:29:43,010 --> 00:29:39,240

languages

721

00:29:46,730 --> 00:29:43,020

no we tested it on a smaller subset I

722

00:29:50,389 --> 00:29:48,710

um is it just computationally too

723

00:29:52,549 --> 00:29:50,399

difficult to do all of them like it

724

00:29:54,710 --> 00:29:52,559

takes up too much time can you not just

725

00:29:56,750 --> 00:29:54,720

tell the computer to run with it

726

00:29:58,909 --> 00:29:56,760

the problem is that you need to build

727

00:30:00,169 --> 00:29:58,919

What's called the language model for

728

00:30:03,529 --> 00:30:00,179

each language

729

00:30:06,130 --> 00:30:03,539

and for that you need a lot of texts and

730

00:30:08,810 --> 00:30:06,140

the European languages

731

00:30:11,570 --> 00:30:08,820

usually all of them have a lot of text

732

00:30:14,149 --> 00:30:11,580

like people write newspapers in them but

733

00:30:16,850 --> 00:30:14,159

if you pick languages that are very

734

00:30:19,430 --> 00:30:16,860

small or very exotic and it's very

735

00:30:21,970 --> 00:30:19,440

difficult to find any electronic text

736

00:30:25,970 --> 00:30:21,980

written in those languages

737

00:30:28,370 --> 00:30:25,980

so so then it's very difficult to

738

00:30:30,409 --> 00:30:28,380

to derive a language model from those

739

00:30:33,350 --> 00:30:30,419

texts because they are too small

740

00:30:36,470 --> 00:30:33,360

that was the reason

741

00:30:38,269 --> 00:30:36,480

yeah it's quite the conundrum

742

00:30:40,190 --> 00:30:38,279

at least you were able to develop some

743

00:30:42,470 --> 00:30:40,200

new techniques that can be applied to

744

00:30:43,730 --> 00:30:42,480

other problems have you made any other

745

00:30:46,130 --> 00:30:43,740

progress other than what you've just

746

00:30:49,990 --> 00:30:46,140

indicated

747

00:30:52,010 --> 00:30:50,000

so I mentioned that we worked on another

748

00:30:53,570 --> 00:30:52,020

undeciphered text how about we just

749

00:30:55,430 --> 00:30:53,580

transition to that and I'll come back

750

00:30:56,990 --> 00:30:55,440

and forth to the Voynich at different

751
00:30:58,970 --> 00:30:57,000
points why don't you tell us about the

752
00:31:01,730 --> 00:30:58,980
Dora Bella cipherite

753
00:31:03,470 --> 00:31:01,740
you know for me like I know there are

754
00:31:06,049 --> 00:31:03,480
people that just spend their all their

755
00:31:08,090 --> 00:31:06,059
lives on Vonage right they're like

756
00:31:10,130 --> 00:31:08,100
obsessed with Voynich but for me it was

757
00:31:12,649 --> 00:31:10,140
just one projects of many

758
00:31:14,870 --> 00:31:12,659
so after running chapter we decided that

759
00:31:17,990 --> 00:31:14,880
we've done everything we could with it

760
00:31:21,889 --> 00:31:18,000
we left it to other people to to puzzle

761
00:31:23,630 --> 00:31:21,899
over and there was another uh Cipher

762
00:31:25,549 --> 00:31:23,640
that caught my attention which is called

763
00:31:29,029 --> 00:31:25,559

the durabella cipher

764

00:31:30,649 --> 00:31:29,039

and this was written in 20th century we

765

00:31:35,690 --> 00:31:30,659

know who wrote it

766

00:31:40,430 --> 00:31:35,700

it was an English composer

767

00:31:43,370 --> 00:31:40,440

who wrote a postcard to his friend

768

00:31:46,310 --> 00:31:43,380

and that first part was was the cipher

769

00:31:47,750 --> 00:31:46,320

it included the cipher which is about 80

770

00:31:50,810 --> 00:31:47,760

characters

771

00:31:54,710 --> 00:31:50,820

in a kind of a strange script

772

00:31:56,810 --> 00:31:54,720

and that that postcard survived and was

773

00:32:01,370 --> 00:31:56,820

published after his death

774

00:32:05,870 --> 00:32:03,470

um and nobody has been able to decipher

775

00:32:07,909 --> 00:32:05,880

that short text

776

00:32:09,889 --> 00:32:07,919

so that's the durable a Cypher another

777

00:32:11,810 --> 00:32:09,899

undeciphered

778

00:32:15,769 --> 00:32:11,820

safer

779

00:32:19,190 --> 00:32:15,779

and now our approach was that maybe you

780

00:32:20,269 --> 00:32:19,200

know this is not a text any language

781

00:32:22,909 --> 00:32:20,279

text

782

00:32:26,330 --> 00:32:22,919

maybe this is just music because that

783

00:32:28,190 --> 00:32:26,340

guy was a composer so what happens what

784

00:32:29,269 --> 00:32:28,200

will happen if we try to decipher into

785

00:32:31,669 --> 00:32:29,279

music

786

00:32:35,090 --> 00:32:31,679

so we came up with algorithms and

787

00:32:38,090 --> 00:32:35,100

implemented programs that can take

788

00:32:41,269 --> 00:32:38,100

a short text like that

789

00:32:43,909 --> 00:32:41,279

uh well even I mean not tax but a short

790

00:32:48,130 --> 00:32:43,919

piece of music that is encoded in some

791

00:32:55,970 --> 00:32:51,830

uh and uh and that's what happened and

792

00:32:58,490 --> 00:32:55,980

we we published a paper that

793

00:33:01,190 --> 00:32:58,500

at the end produces a kind of a

794

00:33:02,690 --> 00:33:01,200

reconstruction of a Melody that is our

795

00:33:06,409 --> 00:33:02,700

best guess

796

00:33:07,669 --> 00:33:06,419

at the decipherment of that Cipher into

797

00:33:10,010 --> 00:33:07,679

music

798

00:33:12,230 --> 00:33:10,020

you said A peculiar statement about the

799

00:33:14,570 --> 00:33:12,240

Voynich manuscript that it may not be a

800

00:33:16,190 --> 00:33:14,580

human language now do you mean to say a

801
00:33:18,409 --> 00:33:16,200
language that large groups of people

802
00:33:21,470 --> 00:33:18,419
speak or that is an alien language like

803
00:33:24,769 --> 00:33:21,480
it's not a homo sapien

804
00:33:26,149 --> 00:33:24,779
it could be a made-up language right so

805
00:33:28,970 --> 00:33:26,159
you know that

806
00:33:31,669 --> 00:33:28,980
actually there exists languages that

807
00:33:35,210 --> 00:33:31,679
were invented like Esperanto

808
00:33:37,250 --> 00:33:35,220
and many languages like hundreds of

809
00:33:39,950 --> 00:33:37,260
languages have been invented this could

810
00:33:42,769 --> 00:33:39,960
be one of those a language that was

811
00:33:44,870 --> 00:33:42,779
never spoken by any Community but

812
00:33:47,570 --> 00:33:44,880
somebody just kind of made up

813
00:33:50,330 --> 00:33:47,580

a language and and it's possible anybody

814

00:33:52,190 --> 00:33:50,340

can do that in learn their own language

815

00:33:54,529 --> 00:33:52,200

I see so it's still it's a human

816

00:33:56,570 --> 00:33:54,539

language in the sense that it's made by

817

00:33:58,490 --> 00:33:56,580

a human but it's not a human language in

818

00:34:00,409 --> 00:33:58,500

the sense that it's not spoken by many

819

00:34:02,149 --> 00:34:00,419

people or even known about

820

00:34:03,409 --> 00:34:02,159

so it's not as if the alligator made up

821

00:34:04,970 --> 00:34:03,419

this language or some other extra

822

00:34:08,089 --> 00:34:04,980

dimensional entity made up the language

823

00:34:10,490 --> 00:34:08,099

or divinely inspired

824

00:34:12,530 --> 00:34:10,500

that's right the better word is probably

825

00:34:14,869 --> 00:34:12,540

natural language we say natural

826

00:34:19,310 --> 00:34:14,879

languages the languages that occur

827

00:34:22,310 --> 00:34:19,320

in in uh on the on the planet spoken by

828

00:34:25,190 --> 00:34:22,320

some community of people I see Stephen

829

00:34:27,050 --> 00:34:25,200

Bax is another professor who is no

830

00:34:29,510 --> 00:34:27,060

longer with us but he studied this

831

00:34:31,849 --> 00:34:29,520

manuscript and I'm curious if you can go

832

00:34:34,790 --> 00:34:31,859

through what his theory is on it are and

833

00:34:38,030 --> 00:34:34,800

then also your commentary on them

834

00:34:41,450 --> 00:34:38,040

well actually I'm not an expert in in

835

00:34:43,629 --> 00:34:41,460

history or or any other theories

836

00:34:46,550 --> 00:34:43,639

you know the the

837

00:34:48,770 --> 00:34:46,560

ultimate test of a theory is that it

838

00:34:50,389 --> 00:34:48,780

produces a decipherment

839

00:34:53,570 --> 00:34:50,399

right so

840

00:34:56,810 --> 00:34:53,580

uh as far as I know no reasonable

841

00:34:58,970 --> 00:34:56,820

decipherment has been produced by Dr Bax

842

00:35:03,050 --> 00:34:58,980

or anybody else

843

00:35:04,970 --> 00:35:03,060

so uh it's it's not a huge motivation to

844

00:35:07,970 --> 00:35:04,980

study somebody's method

845

00:35:10,069 --> 00:35:07,980

if that method has not actually worked

846

00:35:11,569 --> 00:35:10,079

but how does one go about the process of

847

00:35:14,450 --> 00:35:11,579

learning a language from a computational

848

00:35:20,329 --> 00:35:17,510

so you know everybody speaks a language

849

00:35:22,670 --> 00:35:20,339

that's the universal thing every human

850

00:35:24,950 --> 00:35:22,680

being they have their own native

851
00:35:26,270 --> 00:35:24,960
language plus the image speak other

852
00:35:28,970 --> 00:35:26,280
languages

853
00:35:30,710 --> 00:35:28,980
but the majority of people I I think

854
00:35:32,569 --> 00:35:30,720
they just learn very well their own

855
00:35:34,450 --> 00:35:32,579
native language and they learn it as

856
00:35:37,310 --> 00:35:34,460
children

857
00:35:40,130 --> 00:35:37,320
if you try to learn a language

858
00:35:42,470 --> 00:35:40,140
after you're like something somehow like

859
00:35:44,510 --> 00:35:42,480
10 years old then you'll find out that

860
00:35:46,910 --> 00:35:44,520
it actually becomes a different process

861
00:35:48,530 --> 00:35:46,920
it becomes more difficult and you

862
00:35:52,190 --> 00:35:48,540
actually have to school you have to go

863
00:35:53,329 --> 00:35:52,200

to school or study books or go on the

864

00:35:56,569 --> 00:35:53,339

internet

865

00:35:58,730 --> 00:35:56,579

and somebody teaches you a language this

866

00:36:00,650 --> 00:35:58,740

is not how children language

867

00:36:02,270 --> 00:36:00,660

so this is a big difference between the

868

00:36:05,390 --> 00:36:02,280

native language and the second language

869

00:36:06,890 --> 00:36:05,400

that we learn for example when I speak

870

00:36:09,650 --> 00:36:06,900

different you can probably tell that

871

00:36:12,530 --> 00:36:09,660

English is not my first language

872

00:36:14,870 --> 00:36:12,540

my first language is Polish

873

00:36:17,089 --> 00:36:14,880

so

874

00:36:19,730 --> 00:36:17,099

because it's not my native language and

875

00:36:22,609 --> 00:36:19,740

because I learned it as a teenager

876

00:36:26,210 --> 00:36:22,619

you can tell from my accent that uh that

877

00:36:28,390 --> 00:36:26,220

I'm not an native speaker right so uh so

878

00:36:30,890 --> 00:36:28,400

this already tells you something about

879

00:36:33,650 --> 00:36:30,900

what people call the language Instinct

880

00:36:35,089 --> 00:36:33,660

the the ability of people to acquire

881

00:36:40,010 --> 00:36:35,099

language

882

00:36:42,290 --> 00:36:40,020

now Linguistics is is uh is a

883

00:36:45,349 --> 00:36:42,300

science of

884

00:36:47,089 --> 00:36:45,359

of the language which deals with various

885

00:36:49,910 --> 00:36:47,099

aspects of the language and those

886

00:36:51,069 --> 00:36:49,920

include things like formatics and

887

00:36:54,670 --> 00:36:51,079

morphology

888

00:36:57,349 --> 00:36:54,680

grammar syntax semantics pragmatics

889

00:36:59,210 --> 00:36:57,359

acquisition many things what are

890

00:37:02,450 --> 00:36:59,220

pragmatics briefly sorry

891

00:37:04,970 --> 00:37:02,460

yeah pragmatics is probably what you are

892

00:37:08,270 --> 00:37:04,980

most interested yourself is

893

00:37:11,930 --> 00:37:08,280

is uh basically for example a sentiment

894

00:37:13,790 --> 00:37:11,940

analysis is pragmatics right if you if

895

00:37:15,349 --> 00:37:13,800

you deal with sentiment analysis you're

896

00:37:18,349 --> 00:37:15,359

not really interested in finding out

897

00:37:20,750 --> 00:37:18,359

what people say but what they feel about

898

00:37:22,609 --> 00:37:20,760

what they say right so that's that's

899

00:37:25,310 --> 00:37:22,619

what we call pragmatics it's not just

900

00:37:27,829 --> 00:37:25,320

about the message it's about all the

901
00:37:30,109 --> 00:37:27,839
other stuff upon it how do we feel about

902
00:37:32,829 --> 00:37:30,119
the message that sounds terribly

903
00:37:40,550 --> 00:37:37,010
is it no it does well it's it is

904
00:37:42,349 --> 00:37:40,560
difficult but uh but uh but it's doable

905
00:37:44,810 --> 00:37:42,359
and it's not the hardest part it's it is

906
00:37:46,790 --> 00:37:44,820
one of the tasks that people do and and

907
00:37:49,490 --> 00:37:46,800
we have programs now that are very good

908
00:37:50,990 --> 00:37:49,500
at it okay so continue on where you were

909
00:37:53,510 --> 00:37:51,000
please

910
00:37:55,370 --> 00:37:53,520
yeah when you say that it sounds uh very

911
00:37:58,190 --> 00:37:55,380
complicated is because it's hard to

912
00:38:00,650 --> 00:37:58,200
Define exactly what we mean by things

913
00:38:03,410 --> 00:38:00,660

like sentiment like me you mean like

914

00:38:05,870 --> 00:38:03,420

what do you mean sentiment like and then

915

00:38:07,010 --> 00:38:05,880

people say well are you angry are you

916

00:38:10,130 --> 00:38:07,020

happy

917

00:38:12,950 --> 00:38:10,140

are you sad

918

00:38:15,349 --> 00:38:12,960

and then how many feelings do we have

919

00:38:18,290 --> 00:38:15,359

well we have eight feelings

920

00:38:19,970 --> 00:38:18,300

really eight no maybe 12 right so this

921

00:38:22,490 --> 00:38:19,980

these are things that are very difficult

922

00:38:25,550 --> 00:38:22,500

to Define it's much easier to deal with

923

00:38:27,589 --> 00:38:25,560

things like letters or phonemes where we

924

00:38:29,210 --> 00:38:27,599

know exactly how many letters or phonies

925

00:38:32,390 --> 00:38:29,220

we have in the language

926
00:38:35,630 --> 00:38:32,400
and it's easier to write programs that

927
00:38:38,030 --> 00:38:35,640
deal with that yeah so I'll I'll give it

928
00:38:41,390 --> 00:38:38,040
a try so so you can imagine it's like a

929
00:38:44,450 --> 00:38:41,400
pipeline so you start with when you hear

930
00:38:46,069 --> 00:38:44,460
somebody speaking you start with what

931
00:38:47,930 --> 00:38:46,079
are the sounds of the language right

932
00:38:50,630 --> 00:38:47,940
that's phonetics

933
00:38:52,849 --> 00:38:50,640
and now once you've done that then

934
00:38:55,730 --> 00:38:52,859
you've tried to figure out where one

935
00:38:58,250 --> 00:38:55,740
word starts the the other ends right you

936
00:39:00,230 --> 00:38:58,260
want to see you want to identify the

937
00:39:02,270 --> 00:39:00,240
words because there is only a limited

938
00:39:05,930 --> 00:39:02,280

number of words

939

00:39:08,030 --> 00:39:05,940

and that's the what we call lexicon or

940

00:39:10,430 --> 00:39:08,040

lexicos

941

00:39:13,310 --> 00:39:10,440

and then you when you look at the words

942

00:39:15,230 --> 00:39:13,320

you see that they are made up of sounds

943

00:39:17,270 --> 00:39:15,240

or letters

944

00:39:19,089 --> 00:39:17,280

but they also made us something bigger

945

00:39:21,890 --> 00:39:19,099

which is called morphemes

946

00:39:24,349 --> 00:39:21,900

and uh and that's the stuff of

947

00:39:28,730 --> 00:39:24,359

morphology that's the study of

948

00:39:31,069 --> 00:39:28,740

morphology for example if uh if I say a

949

00:39:32,870 --> 00:39:31,079

word like ungrammaticality

950

00:39:35,510 --> 00:39:32,880

and then you can say well there are

951
00:39:37,010 --> 00:39:35,520
three parts of it the an the grammar and

952
00:39:40,130 --> 00:39:37,020
the ality

953
00:39:43,190 --> 00:39:40,140
and that's the morphology

954
00:39:46,370 --> 00:39:43,200
so so these are considered the kind of a

955
00:39:49,190 --> 00:39:46,380
low level low levels of language and as

956
00:39:51,829 --> 00:39:49,200
you go up it becomes more interesting so

957
00:39:53,630 --> 00:39:51,839
first of all how are words put together

958
00:39:56,810 --> 00:39:53,640
into sentences

959
00:39:59,450 --> 00:39:56,820
how is it that you can have sentences

960
00:40:01,250 --> 00:39:59,460
that you ask somebody is that a proper

961
00:40:03,589 --> 00:40:01,260
English sentence and they say yes or no

962
00:40:06,410 --> 00:40:03,599
they can tell even though they have no

963
00:40:08,569 --> 00:40:06,420

idea they haven't studied linguistics

964

00:40:10,849 --> 00:40:08,579

every native speaker can tell you if a

965

00:40:13,609 --> 00:40:10,859

sentence is grammatical or not

966

00:40:15,170 --> 00:40:13,619

that's the study that Noam Chomsky did

967

00:40:17,450 --> 00:40:15,180

in the 50s

968

00:40:18,829 --> 00:40:17,460

can we write a program that can tell a

969

00:40:21,589 --> 00:40:18,839

grammatical sentence from an

970

00:40:24,470 --> 00:40:21,599

ungrammatical sentence

971

00:40:26,270 --> 00:40:24,480

and on top of that on top of syntax is

972

00:40:29,089 --> 00:40:26,280

semantics which is about the meaning of

973

00:40:31,490 --> 00:40:29,099

words we can have perfectly Dramatical

974

00:40:33,650 --> 00:40:31,500

sentences that are meaningless

975

00:40:37,310 --> 00:40:33,660

and vice versa we can have meaningless

976

00:40:39,530 --> 00:40:37,320

utterances that are not grammatical

977

00:40:41,569 --> 00:40:39,540

and then on top of cement it is

978

00:40:44,210 --> 00:40:41,579

pragmatics which has all these things

979

00:40:47,390 --> 00:40:44,220

that are difficult to Define and that

980

00:40:49,790 --> 00:40:47,400

appear to you very complicated this

981

00:40:51,829 --> 00:40:49,800

Universal grammar of chomsky's it's true

982

00:40:53,150 --> 00:40:51,839

in the sense that you can create a

983

00:40:54,650 --> 00:40:53,160

program that can identify which

984

00:40:57,470 --> 00:40:54,660

sentences are grammatically correct and

985

00:41:02,210 --> 00:41:00,230

actually I don't think so I think that's

986

00:41:03,589 --> 00:41:02,220

what Chomsky tried to do all his life

987

00:41:08,150 --> 00:41:03,599

but

988

00:41:09,410 --> 00:41:08,160

it has not it has not been done as far

989

00:41:12,530 --> 00:41:09,420

as I know

990

00:41:15,770 --> 00:41:12,540

but at least that was the state of the

991

00:41:18,589 --> 00:41:15,780

earth about 10 years ago now the the the

992

00:41:21,650 --> 00:41:18,599

last few years we have seen the the

993

00:41:23,810 --> 00:41:21,660

neural language models appearing which

994

00:41:26,089 --> 00:41:23,820

are extremely effective

995

00:41:28,490 --> 00:41:26,099

and which as you know can produce a

996

00:41:31,130 --> 00:41:28,500

completely grammatical and text that

997

00:41:33,710 --> 00:41:31,140

also makes sense yeah so so by extension

998

00:41:35,930 --> 00:41:33,720

that means that these these uh programs

999

00:41:37,910 --> 00:41:35,940

can tell the difference between a

1000

00:41:39,530 --> 00:41:37,920

grammatical and ungrammatical centers

1001
00:41:41,089 --> 00:41:39,540
because they only produce grammatical

1002
00:41:43,490 --> 00:41:41,099
sentence

1003
00:41:49,370 --> 00:41:43,500
are there other Universal Concepts in

1004
00:41:55,670 --> 00:41:52,569
uh so the universal Concepts in language

1005
00:41:56,829 --> 00:41:55,680
are the things that are in every

1006
00:42:02,450 --> 00:41:56,839
language

1007
00:42:04,550 --> 00:42:02,460
there is something that almost all

1008
00:42:06,950 --> 00:42:04,560
languages possess but some languages

1009
00:42:09,290 --> 00:42:06,960
don't then it's not Universal

1010
00:42:12,410 --> 00:42:09,300
there is a whole area of linguistics

1011
00:42:15,230 --> 00:42:12,420
that is dealing with finding things that

1012
00:42:18,230 --> 00:42:15,240
are Universal in human languages

1013
00:42:20,990 --> 00:42:18,240

and I as far as I know there is a long

1014

00:42:22,910 --> 00:42:21,000

list of those things have you used any

1015

00:42:25,190 --> 00:42:22,920

machine learning or neural language

1016

00:42:27,430 --> 00:42:25,200

processing in the decipherment of the

1017

00:42:30,530 --> 00:42:27,440

Voynich

1018

00:42:33,950 --> 00:42:30,540

we did use machine learning but not

1019

00:42:37,130 --> 00:42:33,960

neural methods no the the

1020

00:42:40,069 --> 00:42:37,140

reason we didn't try we didn't use the

1021

00:42:42,170 --> 00:42:40,079

neural methods for decipherment

1022

00:42:44,810 --> 00:42:42,180

I I think you have some experience

1023

00:42:48,710 --> 00:42:44,820

already with these neural Bots is that

1024

00:42:50,750 --> 00:42:48,720

they can make sense of everything so for

1025

00:42:53,450 --> 00:42:50,760

example Google translate if you give it

1026
00:42:55,190 --> 00:42:53,460
a something that doesn't make sense it

1027
00:42:58,250 --> 00:42:55,200
will still translate it

1028
00:42:59,990 --> 00:42:58,260
into something that does obviously we

1029
00:43:02,750 --> 00:43:00,000
don't want something like that to be

1030
00:43:04,750 --> 00:43:02,760
applied to voynish manuscript because we

1031
00:43:08,089 --> 00:43:04,760
want to really know what's really there

1032
00:43:10,309 --> 00:43:08,099
not how to make sense out of it

1033
00:43:12,950 --> 00:43:10,319
in some way right

1034
00:43:14,510 --> 00:43:12,960
there isn't some way of identifying what

1035
00:43:16,670 --> 00:43:14,520
makes sense and what doesn't in the same

1036
00:43:18,050 --> 00:43:16,680
way that for some sentences you can

1037
00:43:20,329 --> 00:43:18,060
identify if it's grammatically correct

1038
00:43:22,069 --> 00:43:20,339

or incorrect like that program has not

1039

00:43:24,230 --> 00:43:22,079

been completely explicated like you

1040

00:43:25,910 --> 00:43:24,240

mentioned with Chomsky but maybe there's

1041

00:43:28,069 --> 00:43:25,920

huge progress there is there not

1042

00:43:29,870 --> 00:43:28,079

progress in saying the sentence makes

1043

00:43:33,170 --> 00:43:29,880

sense or not

1044

00:43:34,910 --> 00:43:33,180

that is much harder to do there is

1045

00:43:35,870 --> 00:43:34,920

progress yeah every year there is

1046

00:43:39,050 --> 00:43:35,880

progress

1047

00:43:40,670 --> 00:43:39,060

but we are still far from reaching that

1048

00:43:44,030 --> 00:43:40,680

point

1049

00:43:47,329 --> 00:43:44,040

you've seen that there's chat GPT and

1050

00:43:48,950 --> 00:43:47,339

there's open AIS gpt3 what's your

1051
00:43:52,730 --> 00:43:48,960
opinion of them are you excited by them

1052
00:43:59,809 --> 00:43:56,270
I am excited that those tools become

1053
00:44:02,030 --> 00:43:59,819
available but I'm also kind of worried

1054
00:44:03,109 --> 00:44:02,040
that people are too enthusiastic about

1055
00:44:06,829 --> 00:44:03,119
them

1056
00:44:10,130 --> 00:44:06,839
and they for me the problem is that

1057
00:44:13,010 --> 00:44:10,140
they are basically what somebody called

1058
00:44:15,230 --> 00:44:13,020
parrots right they they're parrots that

1059
00:44:18,349 --> 00:44:15,240
have heard a lot of language being

1060
00:44:20,750 --> 00:44:18,359
spoken everything that was ever written

1061
00:44:22,849 --> 00:44:20,760
and they are very good at repeating

1062
00:44:24,109 --> 00:44:22,859
putting together those sentences and

1063
00:44:26,450 --> 00:44:24,119

words together

1064

00:44:28,250 --> 00:44:26,460

but there is no really under no real

1065

00:44:32,210 --> 00:44:28,260

understanding underneath

1066

00:44:32,990 --> 00:44:32,220

those systems cannot tell us why they

1067

00:44:35,329 --> 00:44:33,000

think

1068

00:44:37,970 --> 00:44:35,339

the these things that they say are true

1069

00:44:39,829 --> 00:44:37,980

they basically repeating

1070

00:44:42,829 --> 00:44:39,839

the words that have been written

1071

00:44:45,829 --> 00:44:42,839

somewhere and rearranging

1072

00:44:46,970 --> 00:44:45,839

to be fair most people when they're

1073

00:44:48,470 --> 00:44:46,980

putting out something that's creative

1074

00:44:49,790 --> 00:44:48,480

they're just repeating what they've seen

1075

00:44:52,609 --> 00:44:49,800

and they're mixing it up and they

1076
00:44:54,349 --> 00:44:52,619
believe it to be absolutely new and also

1077
00:44:55,849 --> 00:44:54,359
just so you know there is something

1078
00:44:59,270 --> 00:44:55,859
creative about mixing up and then

1079
00:45:01,250 --> 00:44:59,280
presenting it and furthermore most

1080
00:45:02,630 --> 00:45:01,260
people maybe even all of us we don't

1081
00:45:04,430 --> 00:45:02,640
know the motivations like will

1082
00:45:06,230 --> 00:45:04,440
confabulate some reason for why we

1083
00:45:08,630 --> 00:45:06,240
created so and so like that's why the

1084
00:45:10,430 --> 00:45:08,640
whole field of psychoanalysis came about

1085
00:45:13,730 --> 00:45:10,440
because we don't know why we do what we

1086
00:45:15,230 --> 00:45:13,740
do we make up some reason so why does it

1087
00:45:17,809 --> 00:45:15,240
matter that the computer doesn't know

1088
00:45:19,450 --> 00:45:17,819

why it does what it's doing and that

1089

00:45:22,010 --> 00:45:19,460

it's quote unquote

1090

00:45:24,650 --> 00:45:22,020

repeating while mixing let's say mixing

1091

00:45:28,490 --> 00:45:26,450

I don't think it matters if you're

1092

00:45:30,829 --> 00:45:28,500

interested in a computer producing art

1093

00:45:31,730 --> 00:45:30,839

like writing a song or painting a

1094

00:45:34,069 --> 00:45:31,740

picture

1095

00:45:36,230 --> 00:45:34,079

but it doesn't matter if you rely on the

1096

00:45:37,690 --> 00:45:36,240

computer to tell you what the truth is

1097

00:45:41,690 --> 00:45:37,700

right

1098

00:45:43,250 --> 00:45:41,700

because if you don't if somebody cannot

1099

00:45:44,930 --> 00:45:43,260

explain to you why they believe

1100

00:45:48,770 --> 00:45:44,940

something is true

1101

00:45:50,750 --> 00:45:48,780

then how can you trust them

1102

00:45:54,710 --> 00:45:50,760

these are deep questions

1103

00:45:57,050 --> 00:45:54,720

so yeah what I find remarkable is that

1104

00:45:59,870 --> 00:45:57,060

you can just even a simple program

1105

00:46:01,790 --> 00:45:59,880

asking it to code this in Python code

1106

00:46:03,349 --> 00:46:01,800

something that does this in Python code

1107

00:46:06,109 --> 00:46:03,359

something that does this in autohotkey

1108

00:46:09,589 --> 00:46:06,119

or whatever it may be and it does it or

1109

00:46:11,690 --> 00:46:09,599

does it 90 the way there who the heck I

1110

00:46:13,609 --> 00:46:11,700

didn't think that that would be possible

1111

00:46:15,470 --> 00:46:13,619

for quite some time there's something

1112

00:46:17,089 --> 00:46:15,480

truthful about that in the sense that it

1113

00:46:19,069 --> 00:46:17,099

works like you can actually test if the

1114

00:46:21,530 --> 00:46:19,079

code works so that's a test of Truth

1115

00:46:23,089 --> 00:46:21,540

more so than a statement are you happy

1116

00:46:24,530 --> 00:46:23,099

about that or you feel like even that

1117

00:46:28,010 --> 00:46:24,540

old technology that could have been done

1118

00:46:34,790 --> 00:46:32,510

so well the program programming is a bit

1119

00:46:37,790 --> 00:46:34,800

different story right because

1120

00:46:40,390 --> 00:46:37,800

you can actually test programs right

1121

00:46:43,309 --> 00:46:40,400

so if some if you ask

1122

00:46:45,050 --> 00:46:43,319

whether it's a human or it's a bot to

1123

00:46:47,390 --> 00:46:45,060

write a program you can

1124

00:46:50,150 --> 00:46:47,400

you provide a specification then you can

1125

00:46:51,829 --> 00:46:50,160

go through the testing the test

1126
00:46:54,470 --> 00:46:51,839
procedure and find out if that program

1127
00:46:56,650 --> 00:46:54,480
really does what it does right so we

1128
00:47:00,470 --> 00:46:56,660
don't actually have to trust

1129
00:47:03,230 --> 00:47:00,480
the pro the trust anything we can just

1130
00:47:05,569 --> 00:47:03,240
test it right but if we don't if we

1131
00:47:07,250 --> 00:47:05,579
don't if if we don't have time to test

1132
00:47:10,370 --> 00:47:07,260
it then I would be wondering whether

1133
00:47:12,589 --> 00:47:10,380
it's a good idea to depend on such a

1134
00:47:15,109 --> 00:47:12,599
program

1135
00:47:18,170 --> 00:47:15,119
so going back to the voyage have you

1136
00:47:20,150 --> 00:47:18,180
thought about if it's composed of at

1137
00:47:22,990 --> 00:47:20,160
least one language like maybe there are

1138
00:47:30,349 --> 00:47:27,230

you know uh it could be it could be a

1139

00:47:34,190 --> 00:47:30,359

lot of things there you know the the you

1140

00:47:36,290 --> 00:47:34,200

can make these uh encryption systems as

1141

00:47:38,990 --> 00:47:36,300

complicated as you wish

1142

00:47:41,089 --> 00:47:39,000

so so it's all possible there is no

1143

00:47:42,950 --> 00:47:41,099

there is no limit there will be no limit

1144

00:47:45,470 --> 00:47:42,960

where we can say well we tried

1145

00:47:48,530 --> 00:47:45,480

everything and now we know it doesn't

1146

00:47:53,569 --> 00:47:48,540

make sense so it must be some kind of a

1147

00:47:56,990 --> 00:47:53,579

joke or some kind of random generator

1148

00:48:01,010 --> 00:47:57,000

but uh what is fascinating about Voynich

1149

00:48:04,630 --> 00:48:01,020

is that we can use it to actually

1150

00:48:08,750 --> 00:48:04,640

create new things right so

1151
00:48:09,950 --> 00:48:08,760
we take decipher and we create a Melody

1152
00:48:13,309 --> 00:48:09,960
right

1153
00:48:16,309 --> 00:48:13,319
and and many people take advantage and

1154
00:48:20,930 --> 00:48:16,319
they produce decipherments that

1155
00:48:23,809 --> 00:48:20,940
are like their own pieces of of art like

1156
00:48:25,930 --> 00:48:23,819
their own books the only problem is that

1157
00:48:29,450 --> 00:48:25,940
everybody produces a different one so

1158
00:48:32,089 --> 00:48:29,460
none of them can be actually correct but

1159
00:48:34,309 --> 00:48:32,099
it is still a creation

1160
00:48:37,130 --> 00:48:34,319
so I think that is that is very good

1161
00:48:39,470 --> 00:48:37,140
about varnish that it exists

1162
00:48:41,930 --> 00:48:39,480
have you thought about Voynich from less

1163
00:48:44,809 --> 00:48:41,940

of a computational perspective and more

1164

00:48:46,309 --> 00:48:44,819

just from a human motivation one what

1165

00:48:49,130 --> 00:48:46,319

the heck is this about why would someone

1166

00:48:50,990 --> 00:48:49,140

go through such lengths to decipher this

1167

00:48:52,190 --> 00:48:51,000

or maybe it's not even lengths like you

1168

00:48:53,990 --> 00:48:52,200

mentioned it could be something trivial

1169

00:48:57,410 --> 00:48:54,000

we're just overlooking like what are

1170

00:49:00,470 --> 00:48:57,420

their theories come up in your mind just

1171

00:49:03,650 --> 00:49:00,480

surmising just conjecture

1172

00:49:06,109 --> 00:49:03,660

yeah so one of of the more interesting

1173

00:49:09,829 --> 00:49:06,119

theories that I've encountered actually

1174

00:49:12,010 --> 00:49:09,839

comes from this U.S expert

1175

00:49:15,050 --> 00:49:12,020

on decipherment

1176

00:49:18,710 --> 00:49:15,060

at in the end he said

1177

00:49:21,890 --> 00:49:18,720

that he thinks this is a an artificial

1178

00:49:23,270 --> 00:49:21,900

language somebody created an artificial

1179

00:49:29,870 --> 00:49:23,280

language

1180

00:49:31,370 --> 00:49:29,880

well if that's the case then it's it's

1181

00:49:33,109 --> 00:49:31,380

very difficult it would be very

1182

00:49:35,809 --> 00:49:33,119

difficult to decipher it because we

1183

00:49:37,010 --> 00:49:35,819

don't know the principles of that

1184

00:49:41,210 --> 00:49:37,020

language

1185

00:49:43,790 --> 00:49:41,220

completely unpronounceable it's just a

1186

00:49:46,250 --> 00:49:43,800

sequence of symbols

1187

00:49:50,329 --> 00:49:46,260

um yeah so I think that's that that is

1188

00:49:53,930 --> 00:49:50,339

one theory that that is for me the most

1189

00:49:58,190 --> 00:49:56,809

what else have you heard that is at

1190

00:50:04,250 --> 00:49:58,200

least somewhat convincing maybe this

1191

00:50:09,530 --> 00:50:06,170

um

1192

00:50:11,089 --> 00:50:09,540

you know when I look anybody can look at

1193

00:50:14,329 --> 00:50:11,099

those illustrations they are on the web

1194

00:50:15,770 --> 00:50:14,339

right and if if you look for them for a

1195

00:50:19,370 --> 00:50:15,780

long time

1196

00:50:21,470 --> 00:50:19,380

sometimes I think this this somebody was

1197

00:50:24,349 --> 00:50:21,480

not quite

1198

00:50:26,809 --> 00:50:24,359

it wasn't it was a work of an expert it

1199

00:50:28,849 --> 00:50:26,819

was a work of somebody who actually

1200

00:50:30,650 --> 00:50:28,859

didn't know what they were doing and

1201

00:50:33,470 --> 00:50:30,660

just try to

1202

00:50:36,410 --> 00:50:33,480

create something like what they saw

1203

00:50:38,089 --> 00:50:36,420

before in other codices in other books

1204

00:50:40,010 --> 00:50:38,099

a little bit like a neural language

1205

00:50:41,510 --> 00:50:40,020

model that just looks a lot of things

1206

00:50:43,550 --> 00:50:41,520

yeah

1207

00:50:46,130 --> 00:50:43,560

she's a lot of reads a lot and then

1208

00:50:49,190 --> 00:50:46,140

produces something that looks like

1209

00:50:51,109 --> 00:50:49,200

it should make sense but it doesn't huh

1210

00:50:52,970 --> 00:50:51,119

that's interesting

1211

00:50:55,250 --> 00:50:52,980

we know that those language models can

1212

00:50:57,049 --> 00:50:55,260

be tricked to produce texts that just

1213

00:50:59,990 --> 00:50:57,059

seem to make sense about complete

1214

00:51:02,510 --> 00:51:00,000

nonsense right for example why it's good

1215

00:51:04,609 --> 00:51:02,520

to eat crushed glass right

1216

00:51:06,470 --> 00:51:04,619

it will give you all the all the reasons

1217

00:51:07,490 --> 00:51:06,480

for that why it is good to eat first

1218

00:51:09,530 --> 00:51:07,500

class

1219

00:51:11,270 --> 00:51:09,540

when it comes to the doorbell Cipher

1220

00:51:12,589 --> 00:51:11,280

there were some other people who came up

1221

00:51:15,290 --> 00:51:12,599

with decipherments I'm going to read

1222

00:51:18,049 --> 00:51:15,300

some right now starts Larks it's chaotic

1223

00:51:20,569 --> 00:51:18,059

but a cloak obscures my new letters a b

1224

00:51:22,910 --> 00:51:20,579

Alpha Beta that is the Greek letters

1225

00:51:24,650 --> 00:51:22,920

below I own the dark

1226

00:51:26,450 --> 00:51:24,660

and then there's others like why am I

1227

00:51:27,530 --> 00:51:26,460

very sad and so on I'm sure you've heard

1228

00:51:29,329 --> 00:51:27,540

these

1229

00:51:30,770 --> 00:51:29,339

yeah what do you mean yeah okay what do

1230

00:51:32,569 --> 00:51:30,780

you make of these

1231

00:51:35,630 --> 00:51:32,579

I've seen this before but it always

1232

00:51:37,970 --> 00:51:35,640

makes me laugh when I hear it's this for

1233

00:51:39,890 --> 00:51:37,980

me is complete nonsense

1234

00:51:42,349 --> 00:51:39,900

why why

1235

00:51:46,069 --> 00:51:42,359

well it is it is nonsense to imagine

1236

00:51:48,049 --> 00:51:46,079

that uh a distinguished uh English

1237

00:51:52,010 --> 00:51:48,059

composer would write something like that

1238

00:51:55,430 --> 00:51:52,020

to his uh love interest

1239

00:51:57,890 --> 00:51:55,440

if you were to decipher Voynich what

1240

00:52:02,109 --> 00:51:57,900

would be next for you no more ciphers or

1241

00:52:08,510 --> 00:52:05,690

you know I I think

1242

00:52:10,250 --> 00:52:08,520

did this happen I mean varnish has not

1243

00:52:13,190 --> 00:52:10,260

been decrypted but there was a very

1244

00:52:17,150 --> 00:52:13,200

interesting uh deciphering recently of

1245

00:52:19,010 --> 00:52:17,160

uh of actual Cipher which was called the

1246

00:52:20,809 --> 00:52:19,020

Zodiac Cypher I don't know issue right

1247

00:52:23,630 --> 00:52:20,819

right yeah

1248

00:52:26,150 --> 00:52:23,640

and that that is actually correct right

1249

00:52:29,270 --> 00:52:26,160

that that decide for me is not fake

1250

00:52:32,390 --> 00:52:29,280

it is actually a correct decipherment

1251
00:52:34,730 --> 00:52:32,400
so I I would probably ask that that

1252
00:52:36,049 --> 00:52:34,740
person about their feelings like what

1253
00:52:39,470 --> 00:52:36,059
how they feel

1254
00:52:41,630 --> 00:52:39,480
about cracking that Cipher is it like a

1255
00:52:43,430 --> 00:52:41,640
complete Bliss or is it like some kind

1256
00:52:46,130 --> 00:52:43,440
of disappointment that

1257
00:52:48,650 --> 00:52:46,140
and I put so much work into it and then

1258
00:52:51,349 --> 00:52:48,660
I find that this text is actually

1259
00:52:53,930 --> 00:52:51,359
kind of you know not interesting at all

1260
00:52:56,329 --> 00:52:53,940
it's like some kind of deranged mind

1261
00:53:00,650 --> 00:52:56,339
writing it

1262
00:53:03,170 --> 00:53:00,660
so uh yes you know there are one one

1263
00:53:04,790 --> 00:53:03,180

kind of tragedy is if you don't achieve

1264

00:53:07,549 --> 00:53:04,800

your goal and the other tragedy is you

1265

00:53:09,349 --> 00:53:07,559

if you do it you should go

1266

00:53:11,630 --> 00:53:09,359

yeah that's interesting hey let's get

1267

00:53:13,910 --> 00:53:11,640

philosophical here to me that means that

1268

00:53:16,250 --> 00:53:13,920

you have to enjoy the process more than

1269

00:53:18,109 --> 00:53:16,260

the state even though there's some end

1270

00:53:19,730 --> 00:53:18,119

States and that's supposedly driving the

1271

00:53:22,130 --> 00:53:19,740

process you have to fall in love with

1272

00:53:25,730 --> 00:53:22,140

the process because you may if you're

1273

00:53:28,430 --> 00:53:25,740

lucky and maybe unlucky reach that state

1274

00:53:32,569 --> 00:53:28,440

yeah absolutely and this is something

1275

00:53:34,670 --> 00:53:32,579

that I do feel about problems in

1276

00:53:39,410 --> 00:53:34,680

computational linguistics that

1277

00:53:41,809 --> 00:53:39,420

and that I love doing this stuff and uh

1278

00:53:43,690 --> 00:53:41,819

I just would be able to do this you know

1279

00:53:47,270 --> 00:53:43,700

for free

1280

00:53:49,370 --> 00:53:47,280

and if because it's it's such such huge

1281

00:53:53,150 --> 00:53:49,380

fun to do this

1282

00:53:56,210 --> 00:53:53,160

but the Vonage was just one of the

1283

00:53:59,569 --> 00:53:56,220

projects that I got interested in and uh

1284

00:54:02,030 --> 00:53:59,579

I learned from the project I I got some

1285

00:54:04,430 --> 00:54:02,040

experience from that project that I

1286

00:54:07,250 --> 00:54:04,440

think made me a better scientist

1287

00:54:09,710 --> 00:54:07,260

so that I can apply this experience to

1288

00:54:11,690 --> 00:54:09,720

the projects that are not that actually

1289

00:54:14,030 --> 00:54:11,700

do have a solution

1290

00:54:16,069 --> 00:54:14,040

well right now and we are very excited

1291

00:54:18,049 --> 00:54:16,079

to be working on semantics on lexical

1292

00:54:21,770 --> 00:54:18,059

semantics and

1293

00:54:24,950 --> 00:54:21,780

and we are proposing uh we you know we

1294

00:54:27,290 --> 00:54:24,960

are finding things that other scientists

1295

00:54:29,809 --> 00:54:27,300

find Maybe

1296

00:54:33,530 --> 00:54:29,819

uh controversial right

1297

00:54:35,270 --> 00:54:33,540

but the huge the huge advantage of of

1298

00:54:38,390 --> 00:54:35,280

the work that we do is that we can

1299

00:54:40,790 --> 00:54:38,400

actually provide proofs mathematical

1300

00:54:43,010 --> 00:54:40,800

proofs of of what we do

1301

00:54:45,890 --> 00:54:43,020

and this is this gives us the

1302

00:54:47,870 --> 00:54:45,900

satisfaction of of actually being

1303

00:54:50,690 --> 00:54:47,880

certain that we are doing something

1304

00:54:53,510 --> 00:54:50,700

right because we can prove it

1305

00:54:55,609 --> 00:54:53,520

going back to loving the road more than

1306

00:54:57,589 --> 00:54:55,619

where you're going I feel the same with

1307

00:54:59,690 --> 00:54:57,599

this podcast it's about theories of

1308

00:55:01,849 --> 00:54:59,700

everything and the physics sense so my

1309

00:55:04,069 --> 00:55:01,859

background is in mathematical physics

1310

00:55:07,490 --> 00:55:04,079

and a part of me I feel like I'll be

1311

00:55:10,250 --> 00:55:07,500

extremely disappointed if I encounter or

1312

00:55:12,650 --> 00:55:10,260

if we discover as people as scientists

1313

00:55:14,630 --> 00:55:12,660

the theory of everything there is

1314

00:55:16,730 --> 00:55:14,640

something that's terribly fun about

1315

00:55:19,510 --> 00:55:16,740

learning it and investigating so I don't

1316

00:55:23,170 --> 00:55:21,829

I don't think you have to worry about

1317

00:55:26,150 --> 00:55:23,180

that

1318

00:55:28,790 --> 00:55:26,160

personally I think

1319

00:55:30,950 --> 00:55:28,800

you know looking at how the universe is

1320

00:55:32,809 --> 00:55:30,960

constructed I'm pretty sure it has some

1321

00:55:35,569 --> 00:55:32,819

built-in mechanisms

1322

00:55:37,549 --> 00:55:35,579

so that we can never actually figure it

1323

00:55:41,030 --> 00:55:37,559

out completely

1324

00:55:43,430 --> 00:55:41,040

what gives you that intuition

1325

00:55:46,849 --> 00:55:43,440

well you know the the

1326
00:55:49,370 --> 00:55:46,859
yeah yeah you know like this when you

1327
00:55:52,730 --> 00:55:49,380
talk about theories of errors in you you

1328
00:55:56,750 --> 00:55:52,740
obviously talk to physicists so uh and

1329
00:55:58,190 --> 00:55:56,760
that deal with quantum mechanics and

1330
00:56:00,650 --> 00:55:58,200
things like that

1331
00:56:04,490 --> 00:56:00,660
and how there are certain principles

1332
00:56:06,950 --> 00:56:04,500
that we can prove that we'll never know

1333
00:56:10,849 --> 00:56:06,960
the truth right like we'll never know

1334
00:56:13,730 --> 00:56:10,859
where the particular particle is what is

1335
00:56:14,510 --> 00:56:13,740
the its exact location and speed and so

1336
00:56:18,349 --> 00:56:14,520
on

1337
00:56:21,710 --> 00:56:18,359
and this uh this just this is for me an

1338
00:56:24,710 --> 00:56:21,720

indication that these things are

1339

00:56:30,530 --> 00:56:24,720

constructed in such a way that we will

1340

00:56:34,270 --> 00:56:32,510

all right well that's hopeful but also

1341

00:56:36,410 --> 00:56:34,280

dismaying

1342

00:56:40,010 --> 00:56:36,420

at least it's both and not one without

1343

00:56:42,470 --> 00:56:40,020

the other so about the zodiac

1344

00:56:44,510 --> 00:56:42,480

if that's a substitution Cipher was it a

1345

00:56:45,950 --> 00:56:44,520

substitution and transposition or just

1346

00:56:50,630 --> 00:56:45,960

substitution

1347

00:56:53,450 --> 00:56:50,640

position and it was a very tricky dance

1348

00:56:55,190 --> 00:56:53,460

position too yeah why is that and would

1349

00:56:58,809 --> 00:56:55,200

your method have worked on the Zodiac

1350

00:57:04,670 --> 00:57:01,849

no the the battle would not work on

1351
00:57:06,829 --> 00:57:04,680
zodiac because the Assumption of our

1352
00:57:09,290 --> 00:57:06,839
methods was that we know where the words

1353
00:57:10,609 --> 00:57:09,300
are so in varnish there are spaces

1354
00:57:12,829 --> 00:57:10,619
between words

1355
00:57:15,230 --> 00:57:12,839
and we made this assumption that this is

1356
00:57:17,030 --> 00:57:15,240
not just to confuse but they are

1357
00:57:19,609 --> 00:57:17,040
actually words right

1358
00:57:22,069 --> 00:57:19,619
now in the Zodiac Cipher there was no

1359
00:57:23,390 --> 00:57:22,079
spaces between words

1360
00:57:25,910 --> 00:57:23,400
so

1361
00:57:28,609 --> 00:57:25,920
although it is possible to kind of

1362
00:57:30,530 --> 00:57:28,619
hypothesis where the spaces are that

1363
00:57:33,470 --> 00:57:30,540

method that particular method would not

1364

00:57:36,410 --> 00:57:33,480

work on zodiac

1365

00:57:40,069 --> 00:57:36,420

and the method used to crack the Zodiac

1366

00:57:45,470 --> 00:57:40,079

Cipher can that method or methods be

1367

00:57:50,690 --> 00:57:48,349

I actually you know I don't think so I I

1368

00:57:53,630 --> 00:57:50,700

think the the the key of the

1369

00:57:57,049 --> 00:57:53,640

decipherment in that case was just

1370

00:57:59,690 --> 00:57:57,059

finding the a specific pattern right

1371

00:58:02,510 --> 00:57:59,700

of transposition

1372

00:58:05,030 --> 00:58:02,520

so it was not any kind of cool new

1373

00:58:07,430 --> 00:58:05,040

Theory that is general and can be

1374

00:58:08,750 --> 00:58:07,440

applied to various things it was just

1375

00:58:12,349 --> 00:58:08,760

kind of a

1376

00:58:13,309 --> 00:58:12,359

stroke of lack there that like trial and

1377

00:58:15,349 --> 00:58:13,319

error

1378

00:58:17,390 --> 00:58:15,359

it is always a child in there when you

1379

00:58:18,589 --> 00:58:17,400

do actual decipherment but what I mean

1380

00:58:22,309 --> 00:58:18,599

is that

1381

00:58:26,150 --> 00:58:22,319

there is no method behind it that can be

1382

00:58:28,370 --> 00:58:26,160

generalized and applied to other things

1383

00:58:29,990 --> 00:58:28,380

how it's Ai and maybe this is a term

1384

00:58:32,150 --> 00:58:30,000

that you don't want to use but how has

1385

00:58:34,250 --> 00:58:32,160

AI aided your field so instead of saying

1386

00:58:36,109 --> 00:58:34,260

AI then reference the a specific model

1387

00:58:37,849 --> 00:58:36,119

like Gans I've changed my field because

1388

00:58:42,470 --> 00:58:37,859

of social supervised learning in the

1389

00:58:50,809 --> 00:58:47,329

yeah so you cannot uh you cannot can you

1390

00:58:52,809 --> 00:58:50,819

avoid the word neural nowadays when you

1391

00:58:54,650 --> 00:58:52,819

talk about language

1392

00:58:58,910 --> 00:58:54,660

understanding

1393

00:59:01,309 --> 00:58:58,920

uh it's it's a it's a powerful uh new

1394

00:59:04,130 --> 00:59:01,319

tool and everybody is very very excited

1395

00:59:06,589 --> 00:59:04,140

about it including myself

1396

00:59:08,230 --> 00:59:06,599

so it of course it changed everything

1397

00:59:11,990 --> 00:59:08,240

because

1398

00:59:14,630 --> 00:59:12,000

what the story of language processing is

1399

00:59:17,089 --> 00:59:14,640

that it started from a kind of a

1400

00:59:19,910 --> 00:59:17,099

symbolic processing

1401
00:59:22,730 --> 00:59:19,920
and then moved into the machine learning

1402
00:59:25,789 --> 00:59:22,740
stage and then evolved into the the

1403
00:59:28,549 --> 00:59:25,799
neural methods which we use nowadays so

1404
00:59:31,309 --> 00:59:28,559
what is exciting about it is that every

1405
00:59:34,010 --> 00:59:31,319
few years you have a new Revolution and

1406
00:59:36,289 --> 00:59:34,020
a new methods and and we make constant

1407
00:59:37,670 --> 00:59:36,299
progress to the point that some people

1408
00:59:39,770 --> 00:59:37,680
think

1409
00:59:42,170 --> 00:59:39,780
that the problem of language has been

1410
00:59:44,210 --> 00:59:42,180
solved but it's not the case

1411
00:59:46,309 --> 00:59:44,220
sorry that the problem of language has

1412
00:59:48,349 --> 00:59:46,319
been solved the program of language

1413
00:59:50,930 --> 00:59:48,359

understanding has been solved that that

1414

00:59:53,510 --> 00:59:50,940

we can basically now have programs that

1415

00:59:55,130 --> 00:59:53,520

will do every language related task that

1416

00:59:57,829 --> 00:59:55,140

we want

1417

01:00:01,190 --> 00:59:57,839

and it's not true who thinks that that's

1418

01:00:06,710 --> 01:00:04,010

well you know when I read these uh

1419

01:00:08,809 --> 01:00:06,720

articles about the the neural Bots that

1420

01:00:13,210 --> 01:00:08,819

can you know write

1421

01:00:16,789 --> 01:00:13,220

uh newspaper articles or or

1422

01:00:18,650 --> 01:00:16,799

compose Symphonies or something that

1423

01:00:21,410 --> 01:00:18,660

sometimes you get an impression that

1424

01:00:23,270 --> 01:00:21,420

well we're done right we can just leave

1425

01:00:25,309 --> 01:00:23,280

it all to the computers

1426

01:00:26,390 --> 01:00:25,319

and they will they will do everything

1427

01:00:30,170 --> 01:00:26,400

for us

1428

01:00:33,829 --> 01:00:30,180

but what I tell my students is that you

1429

01:00:36,230 --> 01:00:33,839

really want to become to be educated to

1430

01:00:37,130 --> 01:00:36,240

be somebody who cannot be replaced by a

1431

01:00:39,530 --> 01:00:37,140

computer

1432

01:00:42,289 --> 01:00:39,540

and I guarantee you that they will never

1433

01:00:49,309 --> 01:00:42,299

be able to replace the most important

1434

01:00:54,829 --> 01:00:51,710

are you sure about that what is it about

1435

01:00:56,630 --> 01:00:54,839

human creativity that a machine can't

1436

01:00:58,789 --> 01:00:56,640

replicate by the way I'm not being

1437

01:01:00,410 --> 01:00:58,799

skeptical I just don't know I'm curious

1438

01:01:02,329 --> 01:01:00,420

what your thoughts are because since

1439

01:01:04,569 --> 01:01:02,339

you're in this field

1440

01:01:08,809 --> 01:01:04,579

well that's exactly what you said

1441

01:01:11,289 --> 01:01:08,819

machine cannot replicate creativity and

1442

01:01:13,609 --> 01:01:11,299

replication are are opposite things

1443

01:01:15,170 --> 01:01:13,619

creativity is doing something that has

1444

01:01:16,849 --> 01:01:15,180

not been done before

1445

01:01:19,789 --> 01:01:16,859

of course you can say well it's just

1446

01:01:23,569 --> 01:01:19,799

kind of building on what was before but

1447

01:01:26,270 --> 01:01:23,579

it's not replicating it's not parroting

1448

01:01:29,930 --> 01:01:26,280

it's creating something new based on a

1449

01:01:34,430 --> 01:01:32,809

well there's this old joke of if you

1450

01:01:36,230 --> 01:01:34,440

want to create an apple pie from scratch

1451

01:01:38,089 --> 01:01:36,240

you have to first create the universe

1452

01:01:39,650 --> 01:01:38,099

it's like well did you get it from the

1453

01:01:41,150 --> 01:01:39,660

farm no I bought it from this okay but

1454

01:01:43,069 --> 01:01:41,160

even if you had it from the farm that

1455

01:01:44,750 --> 01:01:43,079

you grow the dirt did you well yes okay

1456

01:01:45,530 --> 01:01:44,760

but did you make the cow and so on and

1457

01:01:48,890 --> 01:01:45,540

so on

1458

01:01:50,870 --> 01:01:48,900

in a sense whatever we think of as new

1459

01:01:52,549 --> 01:01:50,880

it's so tricky like it depends on what

1460

01:01:55,549 --> 01:01:52,559

what the heck are we defining as novel

1461

01:01:57,289 --> 01:01:55,559

as creative and I'm sure if we could

1462

01:01:58,910 --> 01:01:57,299

look into our brain with a certain

1463

01:02:00,470 --> 01:01:58,920

amount of resolution and we had the

1464

01:02:02,809 --> 01:02:00,480

correct model if it even could be

1465

01:02:04,490 --> 01:02:02,819

modeled computationally but regardless

1466

01:02:06,349 --> 01:02:04,500

maybe there's some non-computational

1467

01:02:08,990 --> 01:02:06,359

model if it can even be modeled

1468

01:02:11,450 --> 01:02:09,000

quote-unquote model the point is that I

1469

01:02:13,130 --> 01:02:11,460

imagine it's conceivable to me that what

1470

01:02:15,170 --> 01:02:13,140

we think of as outputting something

1471

01:02:17,329 --> 01:02:15,180

creative is something that is

1472

01:02:19,730 --> 01:02:17,339

algorithmic like I'm not set on this but

1473

01:02:21,890 --> 01:02:19,740

it's conceivable and if that's the case

1474

01:02:23,630 --> 01:02:21,900

then I don't see why a computer can't do

1475

01:02:25,609 --> 01:02:23,640

it now whether or not a computer can

1476

01:02:27,890 --> 01:02:25,619

feel and understand what it's doing like

1477

01:02:30,230 --> 01:02:27,900

that's a separate problem but the actual

1478

01:02:31,970 --> 01:02:30,240

output I don't see an in principle

1479

01:02:33,650 --> 01:02:31,980

reason why it can't be done and I'm

1480

01:02:35,210 --> 01:02:33,660

telling you this as a romantic like I

1481

01:02:37,010 --> 01:02:35,220

don't want this to be done but I see

1482

01:02:39,170 --> 01:02:37,020

more and more like aspects that we

1483

01:02:41,270 --> 01:02:39,180

thought computers could not do

1484

01:02:43,010 --> 01:02:41,280

it couldn't beat us a chest and then but

1485

01:02:45,410 --> 01:02:43,020

well I can't create art oh my gosh can

1486

01:02:47,750 --> 01:02:45,420

it create art and can't compose music oh

1487

01:02:50,510 --> 01:02:47,760

my gosh can it confuse different musical

1488

01:02:52,010 --> 01:02:50,520

styles so it keeps encroaching

1489

01:02:53,870 --> 01:02:52,020

encroaching into these areas that we

1490

01:02:55,849 --> 01:02:53,880

thought this were just exclusive to

1491

01:02:57,470 --> 01:02:55,859

humans well what's left for us is like

1492

01:02:59,390 --> 01:02:57,480

physical dexterity and that seems to be

1493

01:03:01,670 --> 01:02:59,400

it so far my question is are you

1494

01:03:03,530 --> 01:03:01,680

confident in the statement that there's

1495

01:03:05,690 --> 01:03:03,540

something special about human creativity

1496

01:03:08,510 --> 01:03:05,700

I would like that to be the cake I want

1497

01:03:12,309 --> 01:03:08,520

to be convinced of that

1498

01:03:15,170 --> 01:03:12,319

well first of all when you said uh

1499

01:03:17,210 --> 01:03:15,180

that something cannot be done you cannot

1500

01:03:20,089 --> 01:03:17,220

demonstrate that something cannot be

1501

01:03:22,910 --> 01:03:20,099

done you can only demonstrate that

1502

01:03:27,410 --> 01:03:22,920

something can be done by doing it right

1503

01:03:29,750 --> 01:03:27,420

so I I will not be able or I don't think

1504

01:03:33,589 --> 01:03:29,760

anybody would be able to demonstrate

1505

01:03:37,490 --> 01:03:33,599

that computers cannot do something

1506

01:03:40,490 --> 01:03:37,500

but I am a computer scientist I l've you

1507

01:03:43,190 --> 01:03:40,500

know I programmed a lot

1508

01:03:47,049 --> 01:03:43,200

I work with computers a lot and I know

1509

01:03:49,910 --> 01:03:47,059

that the computers are good at doing

1510

01:03:52,250 --> 01:03:49,920

repeatedly certain things

1511

01:03:54,289 --> 01:03:52,260

and repeating patterns that already

1512

01:03:56,870 --> 01:03:54,299

exist right

1513

01:03:59,450 --> 01:03:56,880

you you cannot have an algorithm to

1514

01:04:00,170 --> 01:03:59,460

create something that does not exist I

1515

01:04:02,809 --> 01:04:00,180

mean

1516

01:04:05,510 --> 01:04:02,819

there is novel that is Meaningful of

1517

01:04:07,490 --> 01:04:05,520

course you can create novel things you

1518

01:04:10,549 --> 01:04:07,500

can create chaos right

1519

01:04:12,410 --> 01:04:10,559

you can create a random generator then

1520

01:04:15,289 --> 01:04:12,420

this sequence of randomly generated

1521

01:04:18,829 --> 01:04:15,299

numbers is unique is it novel

1522

01:04:22,069 --> 01:04:18,839

no because it doesn't make sense

1523

01:04:24,650 --> 01:04:22,079

are you afraid of where AI may be or are

1524

01:04:27,470 --> 01:04:24,660

you more hopeful

1525

01:04:30,109 --> 01:04:27,480

I think it's a serious issue and we have

1526

01:04:32,630 --> 01:04:30,119

to think about it you know because

1527

01:04:35,510 --> 01:04:32,640

the danger I see is that people will

1528

01:04:39,710 --> 01:04:35,520

trust those programs too much

1529

01:04:41,930 --> 01:04:39,720

and they uh we build them and we are

1530

01:04:43,430 --> 01:04:41,940

responsible for telling them what we

1531

01:04:46,190 --> 01:04:43,440

want them to do

1532

01:04:48,829 --> 01:04:46,200

if we don't do this right they may do

1533

01:04:50,630 --> 01:04:48,839

surprising things that we never actually

1534

01:04:53,270 --> 01:04:50,640

anticipated

1535

01:04:56,450 --> 01:04:53,280

I think the key thing is that we want

1536

01:04:58,910 --> 01:04:56,460

these things to be transparent we want

1537

01:05:02,089 --> 01:04:58,920

to know if they tell us

1538

01:05:04,370 --> 01:05:02,099

a statement then we want to know why

1539

01:05:07,370 --> 01:05:04,380

they think the statement is true

1540

01:05:09,230 --> 01:05:07,380

we want them to provide a proof of

1541

01:05:12,410 --> 01:05:09,240

something that they state

1542

01:05:13,609 --> 01:05:12,420

obviously they are not at this level yet

1543

01:05:15,349 --> 01:05:13,619

right

1544

01:05:18,950 --> 01:05:15,359

for example

1545

01:05:20,450 --> 01:05:18,960

they can write basically history books

1546

01:05:22,370 --> 01:05:20,460

right let's

1547

01:05:25,010 --> 01:05:22,380

we don't know whether they are

1548

01:05:27,190 --> 01:05:25,020

hallucinating or is it actual facts

1549

01:05:29,630 --> 01:05:27,200

they're talking about

1550

01:05:31,430 --> 01:05:29,640

so there must be some way of them

1551
01:05:34,549 --> 01:05:31,440
providing evidence

1552
01:05:36,829 --> 01:05:34,559
of what they are saying is true

1553
01:05:40,609 --> 01:05:36,839
like put references when you make a

1554
01:05:42,349 --> 01:05:40,619
statement like exactly exactly yeah so

1555
01:05:43,970 --> 01:05:42,359
we I've been talking to students

1556
01:05:46,130 --> 01:05:43,980
recently

1557
01:05:50,030 --> 01:05:46,140
about

1558
01:05:53,450 --> 01:05:50,040
what is true right like if

1559
01:05:56,349 --> 01:05:53,460
how can we decide if if a sentence is

1560
01:05:59,450 --> 01:05:56,359
true or false me

1561
01:06:01,849 --> 01:05:59,460
and and the fact is that

1562
01:06:04,190 --> 01:06:01,859
you know some people say Everything's

1563
01:06:05,630 --> 01:06:04,200

Relative you know like some people think

1564

01:06:08,630 --> 01:06:05,640

this is true and the other people you

1565

01:06:12,230 --> 01:06:08,640

think this is true I what I what I want

1566

01:06:15,410 --> 01:06:12,240

to the students to do is to decide

1567

01:06:17,510 --> 01:06:15,420

first what is the speaker what's the

1568

01:06:20,569 --> 01:06:17,520

author of the utterance if they think

1569

01:06:23,109 --> 01:06:20,579

it's true or not and this is non-trivial

1570

01:06:25,970 --> 01:06:23,119

but if they can establish that the the

1571

01:06:27,950 --> 01:06:25,980

author of the utterance or sentence

1572

01:06:30,650 --> 01:06:27,960

believes it's true

1573

01:06:32,990 --> 01:06:30,660

then it is true with respect to that

1574

01:06:35,109 --> 01:06:33,000

person right so we can say this is a

1575

01:06:38,930 --> 01:06:35,119

true statement according to this person

1576

01:06:41,750 --> 01:06:38,940

and it it is then kind of clear

1577

01:06:44,930 --> 01:06:41,760

that this is some kind of evidence based

1578

01:06:47,990 --> 01:06:44,940

on somebody's belief

1579

01:06:51,289 --> 01:06:48,000

so I do believe we can tell whether a

1580

01:06:54,109 --> 01:06:51,299

statement is true or false modular

1581

01:06:56,089 --> 01:06:54,119

the author of the statement

1582

01:06:57,589 --> 01:06:56,099

except in the case of AI like in the

1583

01:07:00,230 --> 01:06:57,599

case of people we can because they have

1584

01:07:02,450 --> 01:07:00,240

intentions but AI is no

1585

01:07:04,069 --> 01:07:02,460

currently no is there a subfield in

1586

01:07:05,990 --> 01:07:04,079

computer science that's dedicated to

1587

01:07:08,029 --> 01:07:06,000

this problem how did the machine come

1588

01:07:09,289 --> 01:07:08,039

upon this decision can it explain the

1589

01:07:11,210 --> 01:07:09,299

reasons

1590

01:07:13,190 --> 01:07:11,220

yeah many people are working on that

1591

01:07:15,170 --> 01:07:13,200

because many people have realized that

1592

01:07:17,510 --> 01:07:15,180

this is what we need in order to be able

1593

01:07:21,049 --> 01:07:17,520

to use those tools

1594

01:07:22,910 --> 01:07:21,059

and what's that field called or subfield

1595

01:07:24,950 --> 01:07:22,920

well the field that I worked in is

1596

01:07:28,789 --> 01:07:24,960

called natural language processing

1597

01:07:32,809 --> 01:07:28,799

uh and that's sometimes considered to be

1598

01:07:35,569 --> 01:07:32,819

a synonym of computational linguistics

1599

01:07:38,390 --> 01:07:35,579

is there a name for when you're

1600

01:07:40,910 --> 01:07:38,400

specifically trying to

1601
01:07:42,890 --> 01:07:40,920
pry open that black box and then pull

1602
01:07:45,710 --> 01:07:42,900
out something that is understandable to

1603
01:07:50,529 --> 01:07:48,589
like how did it make the decision

1604
01:07:53,510 --> 01:07:50,539
the word that I've heard uses

1605
01:07:56,329 --> 01:07:53,520
interpretability right so you want to

1606
01:07:59,270 --> 01:07:56,339
have a program that not just does the

1607
01:08:03,049 --> 01:07:59,280
job but is also interpretable so it can

1608
01:08:06,250 --> 01:08:03,059
be we can interpret why it does the job

1609
01:08:10,250 --> 01:08:06,260
as it does so the current

1610
01:08:12,470 --> 01:08:10,260
non-interpretability of AI is that what

1611
01:08:14,690 --> 01:08:12,480
you see as its greatest threat or do you

1612
01:08:16,430 --> 01:08:14,700
see that like you've heard strong AI and

1613
01:08:19,249 --> 01:08:16,440

you've heard of The Singularity and that

1614

01:08:21,470 --> 01:08:19,259

machines may turn on humans or that

1615

01:08:23,269 --> 01:08:21,480

other people may use like if you invert

1616

01:08:26,749 --> 01:08:23,279

certain parameters then a drug that was

1617

01:08:28,970 --> 01:08:26,759

that a machine developed to produce a

1618

01:08:30,530 --> 01:08:28,980

drug that was helpful can be turned to

1619

01:08:32,269 --> 01:08:30,540

produce a drug that's extremely potent

1620

01:08:35,150 --> 01:08:32,279

and deleterious do you see the

1621

01:08:36,890 --> 01:08:35,160

non-interpretability of machines as the

1622

01:08:38,209 --> 01:08:36,900

greatest issue that we have right now or

1623

01:08:39,829 --> 01:08:38,219

is somehow connected to all those other

1624

01:08:41,930 --> 01:08:39,839

issues

1625

01:08:44,209 --> 01:08:41,940

I don't know if it's the greatest issue

1626

01:08:46,490 --> 01:08:44,219

but it's an important issue another

1627

01:08:49,370 --> 01:08:46,500

important issue is the so-called bias

1628

01:08:51,890 --> 01:08:49,380

right the these language models are

1629

01:08:54,349 --> 01:08:51,900

trained on texts that have been written

1630

01:08:57,110 --> 01:08:54,359

by people that are biased right and they

1631

01:08:59,930 --> 01:08:57,120

become biased themselves obviously we

1632

01:09:02,390 --> 01:08:59,940

don't want that to be guided by such

1633

01:09:04,249 --> 01:09:02,400

kind of texts

1634

01:09:06,890 --> 01:09:04,259

there's a phrase that you wrote down

1635

01:09:09,829 --> 01:09:06,900

English orthography is not close to

1636

01:09:11,930 --> 01:09:09,839

Optimal correct can you explain firstly

1637

01:09:14,269 --> 01:09:11,940

what orthography is and then take us

1638

01:09:17,390 --> 01:09:14,279

through that phrase

1639

01:09:21,110 --> 01:09:17,400

yes so orthography is the way we write

1640

01:09:22,490 --> 01:09:21,120

language so English exists primarily as

1641

01:09:25,070 --> 01:09:22,500

the spoken thing

1642

01:09:26,390 --> 01:09:25,080

but we also write it down like as every

1643

01:09:29,090 --> 01:09:26,400

language and

1644

01:09:32,150 --> 01:09:29,100

the orthograph is the way we write down

1645

01:09:35,930 --> 01:09:32,160

the sounds and as you may know English

1646

01:09:38,030 --> 01:09:35,940

doesn't have a very good orthography

1647

01:09:40,729 --> 01:09:38,040

well it doesn't seem to be good because

1648

01:09:42,349 --> 01:09:40,739

it's very hard to learn and people that

1649

01:09:44,570 --> 01:09:42,359

learn English they make a lot of

1650

01:09:47,090 --> 01:09:44,580

spelling errors and even native speakers

1651

01:09:49,370 --> 01:09:47,100

find it difficult to write down words

1652

01:09:54,050 --> 01:09:49,380

that they speak

1653

01:09:57,290 --> 01:09:54,060

so a Noam Chomsky had that a kind of a

1654

01:10:00,530 --> 01:09:57,300

statement that English orthography is

1655

01:10:03,530 --> 01:10:00,540

near optimal right it's close to Optimal

1656

01:10:07,550 --> 01:10:03,540

even though it appears not to be

1657

01:10:09,709 --> 01:10:07,560

so we had the projects when we when we

1658

01:10:12,050 --> 01:10:09,719

kind of showed that

1659

01:10:16,070 --> 01:10:12,060

it actually is not optimal it's not

1660

01:10:26,030 --> 01:10:19,729

and so that's the essence of that paper

1661

01:10:33,590 --> 01:10:29,290

Chomsky had ever had very good reasons

1662

01:10:35,350 --> 01:10:33,600

for saying what what he said but uh you

1663

01:10:39,470 --> 01:10:35,360

know in science

1664

01:10:42,410 --> 01:10:39,480

our job is to question everything right

1665

01:10:46,430 --> 01:10:42,420

and that's what we did in that project

1666

01:10:49,970 --> 01:10:46,440

we we wanted to question that statement

1667

01:10:52,689 --> 01:10:49,980

which seems to be nowadays accepted as

1668

01:10:56,270 --> 01:10:52,699

Truth by everybody

1669

01:10:59,090 --> 01:10:56,280

and to show that to to provide evidence

1670

01:11:00,709 --> 01:10:59,100

for that we we wrote programs and we did

1671

01:11:04,070 --> 01:11:00,719

such simulations

1672

01:11:07,130 --> 01:11:04,080

and we published this to show that

1673

01:11:10,430 --> 01:11:07,140

it is not actually optimal it is it is

1674

01:11:11,450 --> 01:11:10,440

not close to Optima could be much better

1675

01:11:14,450 --> 01:11:11,460

um

1676

01:11:15,229 --> 01:11:14,460

Yeah so basically that that's the point

1677

01:11:17,090 --> 01:11:15,239

here

1678

01:11:18,890 --> 01:11:17,100

what was chomsky's reasons for

1679

01:11:20,870 --> 01:11:18,900

suggesting it was optimal because as you

1680

01:11:23,270 --> 01:11:20,880

pointed out it seems on the face that

1681

01:11:26,090 --> 01:11:23,280

it's clear it's not like the word tough

1682

01:11:28,070 --> 01:11:26,100

is with an F but it ends with GH it

1683

01:11:29,570 --> 01:11:28,080

seems clear that it's not so Chomsky

1684

01:11:31,070 --> 01:11:29,580

must have had some reasons and like you

1685

01:11:32,990 --> 01:11:31,080

mentioned he had good reasons what were

1686

01:11:37,850 --> 01:11:33,000

they and then what was his response to

1687

01:11:44,390 --> 01:11:41,810

yes so Chomsky was when he wrote this in

1688

01:11:47,209 --> 01:11:44,400

the 60s he was going against the the

1689

01:11:49,310 --> 01:11:47,219

consensus right which was that English

1690

01:11:51,530 --> 01:11:49,320

orthography is is

1691

01:11:54,110 --> 01:11:51,540

bad okay

1692

01:11:57,290 --> 01:11:54,120

and he questioned that and he said no

1693

01:11:59,630 --> 01:11:57,300

it's actually near optimal

1694

01:12:01,910 --> 01:11:59,640

and it would take a lot of time to go

1695

01:12:04,310 --> 01:12:01,920

into those arguments which are which are

1696

01:12:06,770 --> 01:12:04,320

reasonable

1697

01:12:09,229 --> 01:12:06,780

however

1698

01:12:12,709 --> 01:12:09,239

every every there is more to it right

1699

01:12:15,350 --> 01:12:12,719

every everything can be interpreted in

1700

01:12:19,570 --> 01:12:15,360

different ways and

1701

01:12:23,510 --> 01:12:19,580

the the main assumption uh that is

1702

01:12:27,290 --> 01:12:23,520

not spoken is that our writing system in

1703

01:12:29,270 --> 01:12:27,300

English is based on a history of English

1704

01:12:33,530 --> 01:12:29,280

and other languages

1705

01:12:36,169 --> 01:12:33,540

for example a lot of English at some

1706

01:12:39,410 --> 01:12:36,179

point was very influenced by French

1707

01:12:41,649 --> 01:12:39,420

about a thousand years ago and and that

1708

01:12:46,430 --> 01:12:41,659

influenced the spelling of English

1709

01:12:48,890 --> 01:12:46,440

now even if we if we could change the

1710

01:12:51,290 --> 01:12:48,900

orthography of English

1711

01:12:53,689 --> 01:12:51,300

to something better if if there is

1712

01:12:56,330 --> 01:12:53,699

something better then that wouldn't be

1713

01:12:59,209 --> 01:12:56,340

practically possible because people are

1714

01:13:00,229 --> 01:12:59,219

just used to the way as it is written

1715

01:13:02,390 --> 01:13:00,239

right now

1716

01:13:04,790 --> 01:13:02,400

and besides English is spoken in many

1717

01:13:06,649 --> 01:13:04,800

different languages in countries and

1718

01:13:09,169 --> 01:13:06,659

those countries would never agree on on

1719

01:13:12,770 --> 01:13:09,179

a new system

1720

01:13:14,330 --> 01:13:12,780

so so in a sense uh Chomsky was right

1721

01:13:15,770 --> 01:13:14,340

about so-called morphological

1722

01:13:19,330 --> 01:13:15,780

consistency

1723

01:13:22,189 --> 01:13:19,340

that words that have the same morphemes

1724

01:13:24,050 --> 01:13:22,199

which are pronounced differently should

1725

01:13:26,750 --> 01:13:24,060

have the same representation for the

1726

01:13:27,890 --> 01:13:26,760

Murphy that representation shouldn't

1727

01:13:29,510 --> 01:13:27,900

change

1728

01:13:31,490 --> 01:13:29,520

but there is also something called

1729

01:13:33,229 --> 01:13:31,500

phonetic consistency and you gave

1730

01:13:36,290 --> 01:13:33,239

example of that

1731

01:13:37,810 --> 01:13:36,300

and that is just not good right there

1732

01:13:40,610 --> 01:13:37,820

are just too many

1733

01:13:43,310 --> 01:13:40,620

arbitrary solutions that reflect the

1734

01:13:44,390 --> 01:13:43,320

pronunciation that as it was 500 years

1735

01:13:46,550 --> 01:13:44,400

ago

1736

01:13:48,470 --> 01:13:46,560

for example they were tough as you said

1737

01:13:51,770 --> 01:13:48,480

was actually pronounced with a consonant

1738

01:13:54,350 --> 01:13:51,780

at the end 500 years ago

1739

01:13:56,750 --> 01:13:54,360

no not tough but there are things like

1740

01:13:58,610 --> 01:13:56,760

though for example which also ends with

1741

01:14:00,890 --> 01:13:58,620

GH there's morphological consistency

1742

01:14:03,890 --> 01:14:00,900

phonetic consistency and then there's

1743

01:14:06,709 --> 01:14:03,900

orthographical optimality can you place

1744

01:14:09,950 --> 01:14:06,719

numbers on those like you can say this

1745

01:14:12,370 --> 01:14:09,960

language is 90 optimal orthographically

1746

01:14:14,689 --> 01:14:12,380

and at 50

1747

01:14:17,209 --> 01:14:14,699

morphologically consistent can you

1748

01:14:20,570 --> 01:14:17,219

actually Place numbers on them

1749

01:14:23,750 --> 01:14:20,580

yeah so let me give it right so for

1750

01:14:25,149 --> 01:14:23,760

example finish is considered an

1751

01:14:28,490 --> 01:14:25,159

extremely good

1752

01:14:33,110 --> 01:14:28,500

orthography is completely consistent you

1753

01:14:34,610 --> 01:14:33,120

know in all uh kind of aspects

1754

01:14:37,870 --> 01:14:34,620

um

1755

01:14:40,189 --> 01:14:37,880

some languages are like

1756

01:14:43,189 --> 01:14:40,199

Croatian for example

1757

01:14:46,910 --> 01:14:43,199

is a the orthography was created under

1758

01:14:49,669 --> 01:14:46,920

the principle right as you speak so that

1759

01:14:52,610 --> 01:14:49,679

has this consistency that

1760

01:14:54,229 --> 01:14:52,620

you can just uh you could you don't you

1761

01:14:55,790 --> 01:14:54,239

never make spelling mistakes you just

1762

01:14:57,770 --> 01:14:55,800

write as you speak

1763

01:14:59,649 --> 01:14:57,780

sorry which language was based like that

1764

01:15:03,169 --> 01:14:59,659

that sounds interesting

1765

01:15:04,910 --> 01:15:03,179

that's uh the the server it used to be

1766

01:15:07,570 --> 01:15:04,920

server creation now these are separate

1767

01:15:11,689 --> 01:15:07,580

languages but it still applies to it

1768

01:15:14,290 --> 01:15:11,699

Croatia Knight now Spanish which many

1769

01:15:18,110 --> 01:15:14,300

people are familiar with this is another

1770

01:15:21,590 --> 01:15:18,120

type of language where you know always

1771

01:15:24,169 --> 01:15:21,600

know how to read something you may you

1772

01:15:27,229 --> 01:15:24,179

may still make spelling mistakes

1773

01:15:29,090 --> 01:15:27,239

but you will never pronounce a written

1774

01:15:31,850 --> 01:15:29,100

word in a wrong way

1775

01:15:34,490 --> 01:15:31,860

so that's another type of consistency

1776

01:15:35,990 --> 01:15:34,500

the English doesn't have either of those

1777

01:15:38,450 --> 01:15:36,000

right you can

1778

01:15:40,250 --> 01:15:38,460

you as a native speaker will probably

1779

01:15:43,070 --> 01:15:40,260

make mistakes unless you have a spell

1780

01:15:46,370 --> 01:15:43,080

checker even though you know perfectly

1781

01:15:49,189 --> 01:15:46,380

well how to pronounce a word

1782

01:15:51,470 --> 01:15:49,199

and me as a second language learner of

1783

01:15:54,530 --> 01:15:51,480

English I will encounter words that I

1784

01:15:57,290 --> 01:15:54,540

just don't know how to pronounce right

1785

01:16:00,050 --> 01:15:57,300

so it is definitely a problem in English

1786

01:16:01,689 --> 01:16:00,060

but other languages are even more

1787

01:16:03,770 --> 01:16:01,699

difficult like the Japanese

1788

01:16:06,290 --> 01:16:03,780

orthographic system is even more

1789

01:16:09,530 --> 01:16:06,300

difficult than English

1790

01:16:12,590 --> 01:16:09,540

I'm curious if English stands out as

1791

01:16:15,350 --> 01:16:12,600

best or worse in some Metric and if so

1792

01:16:18,350 --> 01:16:15,360

which for instance I heard that English

1793

01:16:20,870 --> 01:16:18,360

can convey a complex sentence second

1794

01:16:22,729 --> 01:16:20,880

best something like that and Mandarin is

1795

01:16:24,590 --> 01:16:22,739

first you can think of it as a simple

1796

01:16:26,209 --> 01:16:24,600

language as one that a child may just

1797

01:16:28,250 --> 01:16:26,219

come up with on their own and it goes

1798

01:16:30,470 --> 01:16:28,260

ooh and ah and the complexity of what

1799

01:16:31,850 --> 01:16:30,480

they can convey is small and so somehow

1800

01:16:33,470 --> 01:16:31,860

there's some way of measuring that I

1801
01:16:35,149 --> 01:16:33,480
don't know I don't know the the actual

1802
01:16:36,890 --> 01:16:35,159
terminology I just heard this and I

1803
01:16:38,810 --> 01:16:36,900
heard that English is actually pretty

1804
01:16:40,910 --> 01:16:38,820
great it's second in the world and

1805
01:16:42,229 --> 01:16:40,920
Chinese or sorry or Mandarin is best at

1806
01:16:44,149 --> 01:16:42,239
that but anyway the point is I just

1807
01:16:47,030 --> 01:16:44,159
heard this so what is English great ad

1808
01:16:52,130 --> 01:16:50,149
yes so so English English and Chinese

1809
01:16:53,930 --> 01:16:52,140
have something in common which is that

1810
01:16:57,470 --> 01:16:53,940
they are analytic languages so

1811
01:16:59,930 --> 01:16:57,480
morphology in English is is very basic

1812
01:17:01,430 --> 01:16:59,940
compared to languages like Spanish or

1813
01:17:04,610 --> 01:17:01,440

polish

1814

01:17:06,290 --> 01:17:04,620

oh and in in uh in Chinese it's even

1815

01:17:08,510 --> 01:17:06,300

more simple basically there is no

1816

01:17:11,330 --> 01:17:08,520

morphology at all

1817

01:17:15,110 --> 01:17:11,340

so in that in that sense these

1818

01:17:16,030 --> 01:17:15,120

analytical languages uh reach some kind

1819

01:17:18,310 --> 01:17:16,040

of

1820

01:17:23,689 --> 01:17:18,320

maximum at some

1821

01:17:31,070 --> 01:17:27,530

I I know that um

1822

01:17:32,750 --> 01:17:31,080

English is if you compare things written

1823

01:17:34,970 --> 01:17:32,760

in different languages

1824

01:17:37,850 --> 01:17:34,980

sometimes you see on products like 20

1825

01:17:40,910 --> 01:17:37,860

languages with the same message

1826
01:17:42,649 --> 01:17:40,920
the English text will probably be one of

1827
01:17:44,930 --> 01:17:42,659
the shortest ones

1828
01:17:47,090 --> 01:17:44,940
so I I think this is maybe something

1829
01:17:49,090 --> 01:17:47,100
you're referring to that

1830
01:17:52,790 --> 01:17:49,100
that it can actually

1831
01:17:56,330 --> 01:17:52,800
convey the same message with fewer

1832
01:17:58,070 --> 01:17:56,340
letters or fewer symbols

1833
01:17:59,689 --> 01:17:58,080
reminds me of this joke someone was

1834
01:18:01,370 --> 01:17:59,699
translating I think it's I think this

1835
01:18:03,770 --> 01:18:01,380
actually happened I think it was from

1836
01:18:05,330 --> 01:18:03,780
hideo Kojima who's a video game creator

1837
01:18:07,090 --> 01:18:05,340
and he was on stage she speaks Japanese

1838
01:18:09,709 --> 01:18:07,100

and he says he goes

1839

01:18:12,350 --> 01:18:09,719

It goes on for like 20 seconds 30

1840

01:18:13,370 --> 01:18:12,360

seconds and the translator comes he says

1841

01:18:14,990 --> 01:18:13,380

thank you

1842

01:18:16,669 --> 01:18:15,000

laughs

1843

01:18:17,570 --> 01:18:16,679

you're like that's not what he's like

1844

01:18:19,250 --> 01:18:17,580

just

1845

01:18:20,750 --> 01:18:19,260

if you don't if you're lazy or you've

1846

01:18:23,209 --> 01:18:20,760

forgotten that's fine but just there's

1847

01:18:26,510 --> 01:18:23,219

no way that's all of what he said yeah

1848

01:18:29,090 --> 01:18:26,520

yeah well I I actually lived in Japan

1849

01:18:31,250 --> 01:18:29,100

for a while so this is actually the

1850

01:18:33,709 --> 01:18:31,260

issue of pragmatics right

1851

01:18:36,290 --> 01:18:33,719

uh language human language is not just

1852

01:18:39,050 --> 01:18:36,300

exchanging messages it's there's a lot

1853

01:18:42,950 --> 01:18:39,060

of for example related to politeness

1854

01:18:46,430 --> 01:18:42,960

uh and in in in Japanese you spend a lot

1855

01:18:49,250 --> 01:18:46,440

of time just being polite in addition to

1856

01:18:51,229 --> 01:18:49,260

passing a message ah like sand sand at

1857

01:18:53,510 --> 01:18:51,239

the end of a person's name is that take

1858

01:18:56,030 --> 01:18:53,520

a note I Am lower than you or respect

1859

01:18:58,250 --> 01:18:56,040

yeah there is a lot more tools for

1860

01:18:59,450 --> 01:18:58,260

expressing this kind of relationships in

1861

01:19:02,510 --> 01:18:59,460

Japanese

1862

01:19:05,270 --> 01:19:02,520

uh to give a simpler example in in

1863

01:19:07,790 --> 01:19:05,280

Spanish for example you can refer to

1864

01:19:10,850 --> 01:19:07,800

somebody as two or you

1865

01:19:13,550 --> 01:19:10,860

or usted which is like sir

1866

01:19:16,430 --> 01:19:13,560

but in English it's that that doesn't

1867

01:19:18,830 --> 01:19:16,440

exist you just call you everyone

1868

01:19:20,810 --> 01:19:18,840

so that makes things simpler

1869

01:19:23,169 --> 01:19:20,820

do you know who Larry David is from

1870

01:19:28,910 --> 01:19:23,179

Seinfeld

1871

01:19:30,770 --> 01:19:28,920

David the Creator he said that when

1872

01:19:32,870 --> 01:19:30,780

Caesar was being assassinated by Brutus

1873

01:19:34,729 --> 01:19:32,880

that Brutus said something with the two

1874

01:19:36,290 --> 01:19:34,739

and then then Larry David said that was

1875

01:19:38,510 --> 01:19:36,300

too informal for an assassination you

1876

01:19:40,870 --> 01:19:38,520

should be saying instead

1877

01:19:46,130 --> 01:19:43,910

to end this you did your Master's thesis

1878

01:19:48,709 --> 01:19:46,140

on a theoretical evaluation on selected

1879

01:19:50,870 --> 01:19:48,719

backtracking algorithms so how has your

1880

01:19:53,030 --> 01:19:50,880

perspective on that subject since the

1881

01:19:54,830 --> 01:19:53,040

writing of that thesis changed how is it

1882

01:19:57,830 --> 01:19:54,840

developed

1883

01:20:00,110 --> 01:19:57,840

yeah so I I did this this is part of

1884

01:20:04,610 --> 01:20:00,120

what's called artificial intelligence

1885

01:20:06,110 --> 01:20:04,620

but it's a very formal thing like uh

1886

01:20:08,990 --> 01:20:06,120

constraint called constraint

1887

01:20:11,090 --> 01:20:09,000

satisfaction and what I liked about it

1888

01:20:12,669 --> 01:20:11,100

is that you can actually prove something

1889

01:20:15,470 --> 01:20:12,679

I mean

1890

01:20:17,330 --> 01:20:15,480

unlike in pure Linguistics you can never

1891

01:20:20,290 --> 01:20:17,340

prove anything you can just argue about

1892

01:20:23,390 --> 01:20:20,300

it and then some people will disagree

1893

01:20:25,790 --> 01:20:23,400

but I didn't stay in that area because I

1894

01:20:26,930 --> 01:20:25,800

wanted to work with language that I love

1895

01:20:31,070 --> 01:20:26,940

language

1896

01:20:33,290 --> 01:20:31,080

it's very hard to prove anything because

1897

01:20:36,290 --> 01:20:33,300

there are always exceptions

1898

01:20:37,910 --> 01:20:36,300

but now after though all those years I'm

1899

01:20:40,669 --> 01:20:37,920

coming back to the point that I think

1900

01:20:43,550 --> 01:20:40,679

that I can actually use the language of

1901

01:20:47,090 --> 01:20:43,560

mathematics to describe human language

1902

01:20:49,910 --> 01:20:47,100

and I find this very exciting so I I

1903

01:20:53,750 --> 01:20:49,920

hope to be able to prove things

1904

01:20:56,870 --> 01:20:53,760

and then be actually

1905

01:20:58,850 --> 01:20:56,880

safe in saying that

1906

01:21:00,649 --> 01:20:58,860

I'm I'm saying the right thing I'm

1907

01:21:02,209 --> 01:21:00,659

saying the truth because it has been

1908

01:21:04,130 --> 01:21:02,219

proven

1909

01:21:06,410 --> 01:21:04,140

what's one of the more out there

1910

01:21:07,790 --> 01:21:06,420

theories of the Voynich manuscript as to

1911

01:21:09,530 --> 01:21:07,800

what it's about what it contains

1912

01:21:11,030 --> 01:21:09,540

information on that you don't believe in

1913

01:21:13,310 --> 01:21:11,040

but you find interesting maybe even

1914

01:21:16,070 --> 01:21:13,320

plausible

1915

01:21:17,149 --> 01:21:16,080

so there was this hilarious paper and

1916

01:21:19,729 --> 01:21:17,159

somebody

1917

01:21:22,070 --> 01:21:19,739

trying to show that the language of

1918

01:21:23,810 --> 01:21:22,080

vines is actually Lodge Bank

1919

01:21:28,490 --> 01:21:23,820

I don't know if you've heard about it

1920

01:21:32,030 --> 01:21:28,500

it's it's a it's an invented language

1921

01:21:34,430 --> 01:21:32,040

and and this paper showed to me that you

1922

01:21:37,010 --> 01:21:34,440

can actually uh

1923

01:21:39,649 --> 01:21:37,020

provide evidence for anything for any

1924

01:21:42,290 --> 01:21:39,659

language if it's a large band that was

1925

01:21:44,810 --> 01:21:42,300

invented in the 20th century somebody

1926

01:21:46,729 --> 01:21:44,820

wrote Vonage manuscript in the 15th

1927

01:21:49,370 --> 01:21:46,739

century in that language

1928

01:21:50,750 --> 01:21:49,380

then that means you can basically argue

1929

01:21:54,110 --> 01:21:50,760

for anything

1930

01:21:56,270 --> 01:21:54,120

and that again shows the value of if you

1931

01:21:58,189 --> 01:21:56,280

can actually prove something and in the

1932

01:22:00,350 --> 01:21:58,199

case of the orange manuscript the proof

1933

01:22:01,390 --> 01:22:00,360

would be actually in the pudding which

1934

01:22:04,550 --> 01:22:01,400

means

1935

01:22:07,430 --> 01:22:04,560

deciphering it into some kind of

1936

01:22:09,830 --> 01:22:07,440

text that made sense

1937

01:22:12,530 --> 01:22:09,840

do you think it will be deciphered in

1938

01:22:18,649 --> 01:22:16,310

no I I hope it will be I hope it will

1939

01:22:22,130 --> 01:22:18,659

buy wooden Bet On It

1940

01:22:23,870 --> 01:22:22,140

you know the people said uh before in

1941

01:22:25,970 --> 01:22:23,880

the in the history people often said

1942

01:22:26,990 --> 01:22:25,980

something will never be done and it was

1943

01:22:29,870 --> 01:22:27,000

done

1944

01:22:31,490 --> 01:22:29,880

when I first heard about the Zodiac

1945

01:22:35,090 --> 01:22:31,500

Cipher

1946

01:22:37,669 --> 01:22:35,100

I thought no that's not ever never gonna

1947

01:22:41,030 --> 01:22:37,679

be deciphered because it's probably just

1948

01:22:43,310 --> 01:22:41,040

random night and then it turns out that

1949

01:22:45,950 --> 01:22:43,320

it was deciphered so

1950

01:22:47,990 --> 01:22:45,960

that's a lesson for us

1951

01:22:49,669 --> 01:22:48,000

meaning in the case of the zodiac you

1952

01:22:51,709 --> 01:22:49,679

thought that it was gibberish but he

1953

01:22:53,030 --> 01:22:51,719

didn't actually write anything it's not

1954

01:22:54,770 --> 01:22:53,040

something that was deciphered it's just

1955

01:22:56,930 --> 01:22:54,780

symbols

1956

01:22:58,970 --> 01:22:56,940

yeah I thought it was just the

1957

01:23:01,370 --> 01:22:58,980

intentional gibberish to confuse people

1958

01:23:05,169 --> 01:23:01,380

this is similar to the people that say

1959

01:23:07,610 --> 01:23:05,179

that Vines is a joke right it's uh uh

1960

01:23:10,850 --> 01:23:07,620

they they make the same assumption that

1961

01:23:13,729 --> 01:23:10,860

somebody just did it to confuse people

1962

01:23:16,070 --> 01:23:13,739

well thank you for spending about two

1963

01:23:18,410 --> 01:23:16,080

hours with me or an hour and a half on

1964

01:23:20,090 --> 01:23:18,420

what is potentially a joke but we

1965

01:23:22,610 --> 01:23:20,100

hopefully not

1966

01:23:24,470 --> 01:23:22,620

take care man it's good to speak with

1967

01:23:25,970 --> 01:23:24,480

you thank you it was it was fun talking

1968

01:23:27,470 --> 01:23:25,980

to you

1969

01:23:29,630 --> 01:23:27,480

bye

1970

01:23:31,310 --> 01:23:29,640

the podcast is now concluded thank you

1971

01:23:33,229 --> 01:23:31,320

for watching if you haven't subscribed

1972

01:23:35,209 --> 01:23:33,239

or clicked on that like button now would

1973

01:23:38,030 --> 01:23:35,219

be a great time to do so as each

1974

01:23:40,070 --> 01:23:38,040

subscribe and like helps YouTube push

1975

01:23:41,870 --> 01:23:40,080

this content to more people also I

1976

01:23:44,270 --> 01:23:41,880

recently found out that external links

1977

01:23:46,790 --> 01:23:44,280

count plenty toward the algorithm which

1978

01:23:48,050 --> 01:23:46,800

means that when you share on Twitter on

1979

01:23:50,209 --> 01:23:48,060

Facebook on Reddit

1980

01:23:52,070 --> 01:23:50,219

Etc it shows YouTube that people are

1981

01:23:54,110 --> 01:23:52,080

talking about this outside of YouTube

1982

01:23:55,970 --> 01:23:54,120

which in turn greatly AIDS the

1983

01:23:57,649 --> 01:23:55,980

Distribution on YouTube as well if you'd

1984

01:23:59,950 --> 01:23:57,659

like to support more conversations like

1985

01:24:02,030 --> 01:23:59,960

this then do consider visiting

1986

01:24:04,250 --> 01:24:02,040

theoriesofeverything.org again it's

1987

01:24:06,709 --> 01:24:04,260

support from the sponsors and you that

1988

01:24:08,330 --> 01:24:06,719

allow me to work on toe full-time you

1989

01:24:10,430 --> 01:24:08,340

get early access to ad-free audio

1990

01:24:12,590 --> 01:24:10,440

episodes there as well every dollar

1991

01:24:14,630 --> 01:24:12,600

helps far more than you may think either