ACTIVITY PROBLEMS

Can you change

CORNER MALL

to

CORN MAZE

in just SIX moves?!

Hint: one move = remove, add, or move a letter

1. Remove E 2. Remove R 3. Remove L 4. Remove L 5. Add Z 6. Add E

Source: Blake Allen and Patrick Littell

HOW DOES AUTOCORRECT WORK?

Edit distance is used in computer science to tell how different two pieces of text are.

Each time you remove, add, or move a letter, it adds one to the edit distance.

In the problem you just solved, for example, the edit distance between CORNER MALL and CORN MAZE was 6.

Autocorrect works by identifying misspelled words (by choosing words that don't match the list in its dictionary), then changes them to a similar word – that is, a word with a small edit distance.

Spell check works this way too, though it gives you a list of options to choose from in the order of smallest to largest edit distance.

Autocorrect and spell check need to know more than just edit distance. They also need to know which spelling errors are most likely and which typos are most likely.

Professor Word has invented a machine to read words aloud, but it isn’t working right!

This sign it just read doesn’t make any sense…

“Weiccme to the iccai pcci!

Swimming rules:

1. Nc running
2. Nc spiashing
3. The pcci cicses at 7 pm

Have fun!”

Can you figure out which TWO MISTAKES the machine is making?

1. o → c   2. l → i

HOW DOES TEXT RECOGNITION WORK?

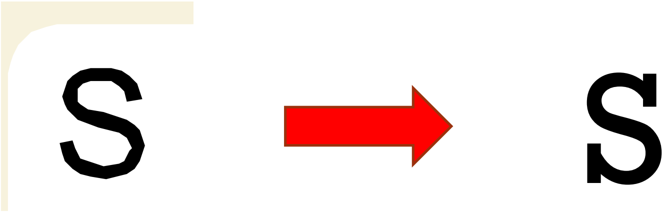
ALSO KNOWN AS OPTICAL CHARACTER RECOGNITION OR OCR

When you type into a computer, the letters are stored as codes in the computer's memory.

When you scan a page into the computer, it creates a picture of the words stored as pixels.

Text recognition was invented to change pixels in into codes for letters.

The computer 'looks' at each black squiggle in the image and tries to match it to one of the letters in its list.



Because a computer can't actually read, though, sometimes it makes mistakes! Unusual fonts and very small letters are a challenge for the program.

The puzzle you just solved is based on real mistakes made by a text recognition program!

DID YOU KNOW?

English is written left-to-right, but not all languages are!

Arabic is written right-to-left:

العربية ←

And Mongolian is written up-to-down!

ᠮᠣᠩᠭᠣᠯ ↓

Here are some sentences in Quechua, a language spoken in Mexico.

Mariya rimashan.	Mary is talking.
Mariya takishan.	Mary is singing.
Mariya kutishan.	Mary is returning.

How would you say “Mary is sitting” in Quechua?

Hint: ‘to sit’ = tiya

Mariya tiyashan.

Here are some sentences in Mapudungun, a language spoken in Chile.

María dungui.	Mary talked.
María petu dungui.	Mary is talking.
Fey ayey.	He laughed.
Fey petu ayey.	He is laughing.
Fey anüi.	She sat.

How would you say "She is sitting" in Mapudungun?

Fey petu anüi.

Here are some sentences in Hindi, a language spoken in India.

Jute laal hain.	The shoes are red
Jute safed hain.	The shoes are white.
Kameez laal hai.	The shirt is red.

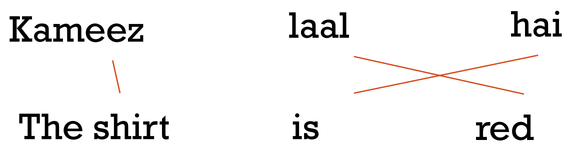
How would you say "The shirt is white" in Hindi?

Kameez safed hai.

HOW DOES COMPUTER TRANSLATION WORK?

The way you solved this puzzle is the same way a computer translator works!

Translation systems are "trained" to notice words and their translations, and the differences in the order of words.



We "train" them by programming them count many millions of word correspondences like the ones in this example. After they count, they compute probabilities.

After training, the computer makes a translation dictionary like this:

kameez = shirt
safed = white
hai = is

The real translation dictionary is full of errors, but it contains a probability for each translation.

When you put in a sentence, the system translates the words using its dictionary and then puts them in the right order. This way, it can translate sentences it has never seen before, without a human having to write them all down!

The shirt is white. →

kameez hai safed

Fill in the blanks so that both phrases make sense!

MOVIE TRAILER PARK

AIR GUITAR HERO

ROCK STAR GAZING

HOW DO COMPUTERS ‘GUESS’ WORDS?

FOR SPEECH RECOGNITION AND COMPUTER TRANSLATION

Sometimes, a computer can’t be sure what word we just said, or which version of a translation is the right one. But how do you teach a machine to guess, and guess well?!

Computers guess words using probabilities. The program ‘looks’ at the words before and after the mystery word, then creates a list of all possible words that could go in that blank. Then, it chooses the one that is the most probable with the word before it, and the word after it:

Looks good!

ROCK	STAR	GAZING

✓

Unfortunately, even when each word in the chain makes sense with the word before it, the whole chain can end up being gibberish! Linguists and computer scientists are working to ‘teach’ computers better ways to guess words, and in the meantime we can enjoy all the funny mistakes computers make.

HOW DO COMPUTERS GUESS WORDS?

FOR SPEECH RECOGNITION AND COMPUTER TRANSLATION

- Let's look at an example from a state-of-the-art computer translation system.
  - Supplied by Austin Matthews.
- The red parts of the sentence are each ok on their own, but together they make a grammatical error:
  - the curriculum will be more emphasis on " real life " problems.
    - Compare to: the curriculum will be more advanced.
  - the curriculum will be more emphasis on " real life " problems.
    - Compare to: the solution will be more emphasis on real life problems.

CAN YOU GUESS THE LANGUAGE?

HINT #1: It's spoken by nearly 65 million people in Southeast Asia.

HINT #2: Its writing system looks like this:

ตัวอักษรไทย

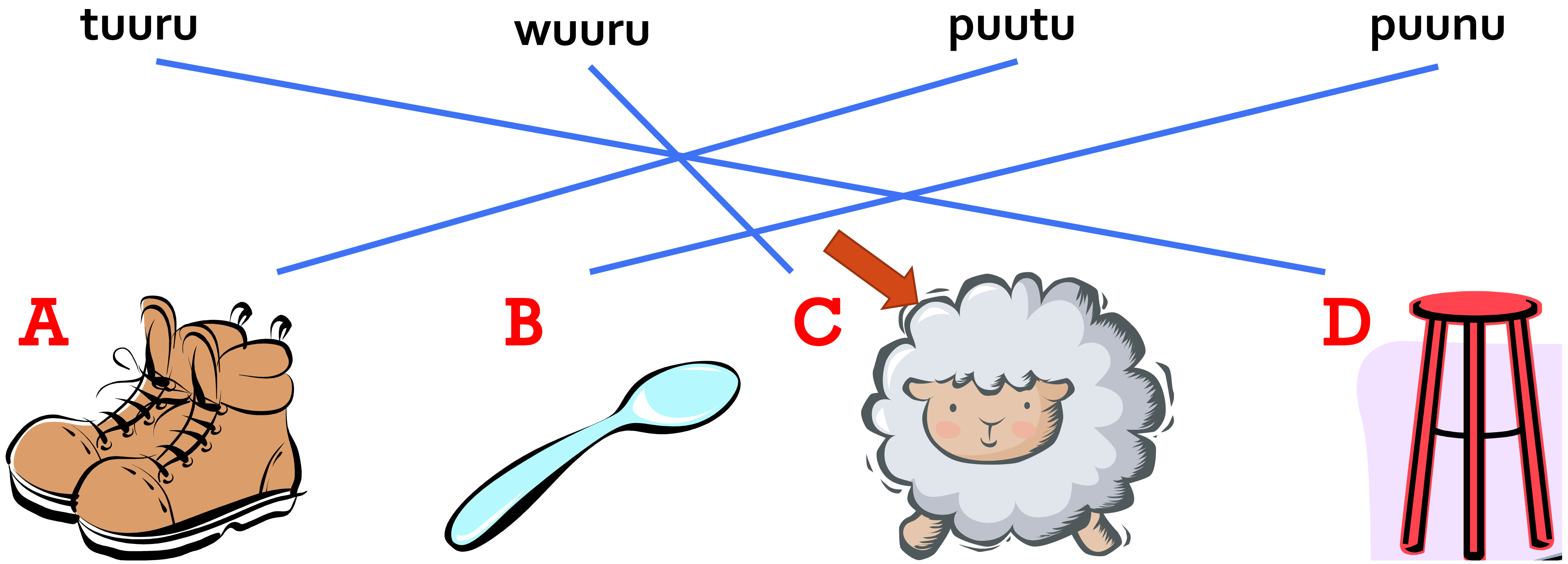
HINT #3: It's closely related to Pali, Sanskrit, Lao, and the minority languages of Thailand.

Thai

Source: www.omniglot.com

Maori is a language spoken by the aboriginal (native) people of New Zealand. Some words in Maori, called loanwords, are "borrowed" from English.

Can you match each loanword to its picture?



tuuru

wuuru

puutu

puunu

Source: Patrick Littell

WHY DO LANGUAGES SOUND DIFFERENT?

Even when a word is the same in English and another language, it might sound very different!

By the time you're six months old, you can already tell the difference between all the sounds of your native language. Not all languages have the same sounds, though!

In this puzzle, you learned Maori speakers don't pronounce the letter 's', and they need to have a vowel after every consonant. So 'stool' becomes 'tuuru'!

Just like the sound 's' is difficult for Maori speakers to pronounce, some sounds might seem unusual to you:

- Nepali speakers use four different kinds of 't'!
- Xhosa has three different clicking sounds that are used as letters!
- Some languages in Central Asia can start a word with four consonants!

Japanese uses a system of letters known as kanji. Each kanji has a specific meaning and pronunciation(s), and kanji can be combined to make new meanings.

日本 = "Japan"

語 = "language"

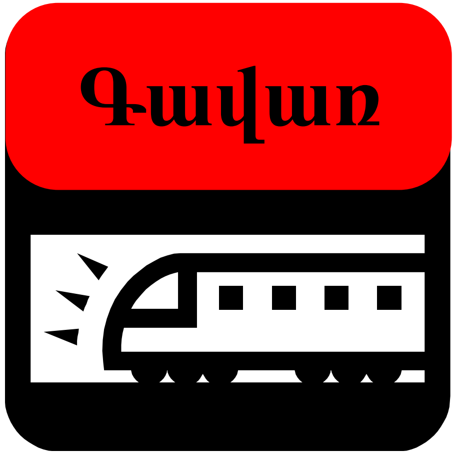
What do you think 日本語 means?

"Japanese"

You are in Kapan in Armenia. You need to get to Ijevan. Can you figure out which way to go just by looking at these Armenian signs?

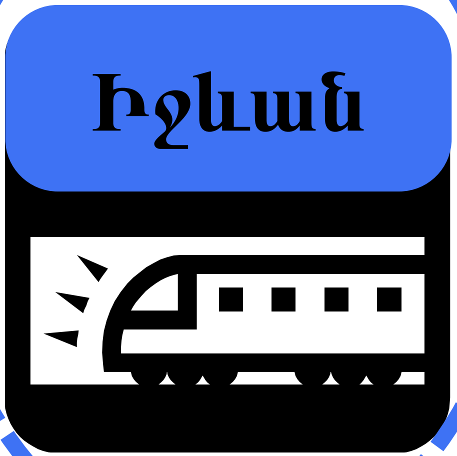
Hint: The sign for Kapan is Կապան in Armenian!

1

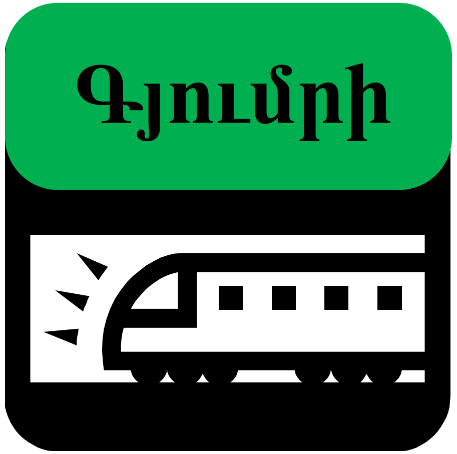


Gavar

2

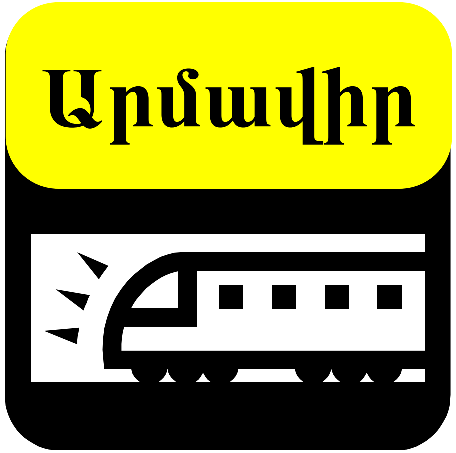


3



Gyumri

4



Armavir

HOW DO OTHER WRITING SYSTEMS WORK?

In these puzzles you saw that some languages aren't written the same way English is.

Linguists divide writing systems into several different categories:

Alphabets use sets of letters to write consonants and vowels: ABCDE

English

Abjads use sets of letters to write only consonants: קומה יהוה ויפצו

Hebrew

Syllabaries use one letter to represent each syllable: すべてのにんげん

Japanese

Bengali

Abugidas combine consonant symbols with vowel symbols to make each letter: সমস্ত মানুষ স্বাধীনভাবে

Semanto-phonetic systems have many letters, each with its own sound and meaning: 漢字

Chinese

Compared to some of these examples, the 26 letters we use to write English i ally small number!

Source: www.omniglot.com

DID YOU KNOW?

Not all languages have the same sounds! Let’s try some sounds not usually found in English.

Glottal stop – the sound in the middle of ‘uh-oh’

Retroflex – press the bottom of your tongue to the roof of your mouth, then let it go while saying ‘t’

Click – press the tip of your tongue to the roof of your mouth, hard, then let go

Each of these newspaper headlines can have two different meanings!

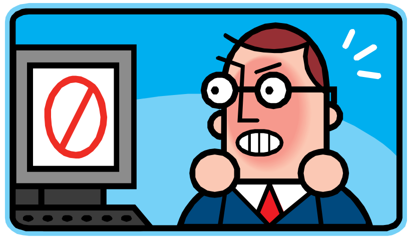
Can you figure out what they are?

A. IRAQI HEAD SEEKS ARMS

B. STOLEN PAINTING FOUND BY TREE

C. KIDS MAKE HEALTHY SNACKS

WHY IS IT SO HARD FOR COMPUTERS TO UNDERSTAND US?



Even though it's funny to imagine the hidden meaning of these sentences, you can probably guess which meaning is correct.

We make thousands of those guesses every day -- we don't always say exactly what we mean, but luckily everyone's brain can fill in the gaps.

Unfortunately, the guesses that are so easy for us are very hard for a computer!
In this example, who is smart and who has computers?

[smart students] and [teachers with computers]

[[smart students] and teachers] with computers

smart [students and [teachers with computers]]

smart [students and teachers] with computers

This made up example is simple compared to what computers really encounter in Wikipedia, social media, email, and on-line newspapers.   Ambiguity of this sort is *combinatoric* and can fill up a computer's memory quickly.

People, on the other hand, read over most ambiguity without noticing. Why doesn't it fill up your memory? That's a good research question!

How many different ways can you break this text into words?

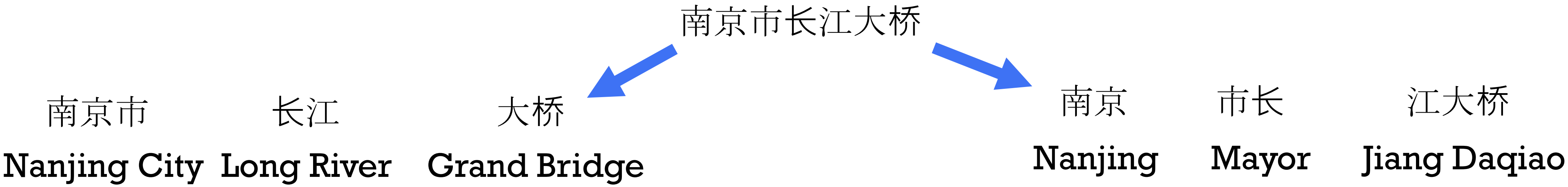
theyouthevent

the youth event
the you the vent
they out he vent

HOW DO COMPUTERS TELL WORDS APART?

In every spoken language there are no spaces between words. Some languages don’t use spaces in writing either.

Chinese writing doesn’t put spaces between words! For example, this Chinese phrase can be broken up two different ways:



南京市长江大桥

南京市 长江 大桥

Nanjing City Long River Grand Bridge

南京 市长 江大桥

Nanjing Mayor Jiang Daqiao

Whether it’s trying to understand spoken English or written Chinese, the computer tells words apart by finding all the possible options, then choosing the one with the highest probability of being correct!

Here are some sentences in Mapudungun, a language spoken in Chile.

chedki	daughter’s son
domo chedki	daughter’s daughter
l̲aku	son’s son
domo l̲aku	son’s daughter

Can you fill in the blank?

Here are some sentences in Inupiaq, an indigenous language spoken in Alaska.

Paniattaaq will not write a book for Aiviq.
Paniattaam maqpiġaaliuġniaŋitkaa Aiviq.

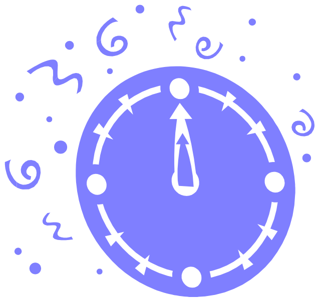
Paniattaaq will write a book for Aiviq.
Paniattaam maqpiġaaliuġniaġaa Aiviq.

Paniattaaq will give Aiviq books.
Paniattaam maqpiġaaksriññiaġaa Aiviq.

How would you say "Paniattaaq will not give Aiviq books?" in Inupiaq?

Paniattaam maqpiġaaksriññiaŋitkaa Aiviq.

Here are some sentences in Japanese.



San ji desu.	It is three o'clock.
Go ji han desu.	It is five thirty.
Roku ji desu.	It is six o'clock.

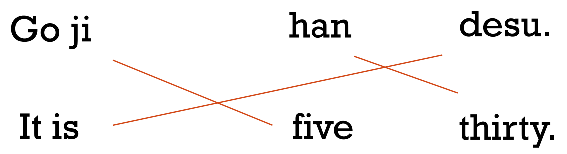
You need to know what time it is, and your friend Erika just told you – in Japanese! Can you figure out what she said?

"San ji han desu." It is three thirty.

HOW DOES COMPUTER TRANSLATION WORK?

The way you solved this puzzle is the same way a computer translator works!

Translation systems are "trained" to notice words and their translations, and the differences in the order of words.



Go ji han desu.

It is five thirty.

We "train" them by programming them count many millions of word correspondences like the ones in this example. After they count, they compute probabilities.

After training, the computer makes a translation dictionary like this:

san = three
ji = o'clock
han = half
desu = is

The real translation dictionary is full of errors, but it contains a probability for each translation.

When you put in a sentence, the system translates the words using its dictionary and then puts them in the right order. This way, it can translate sentences it has never seen before, without a human having to write them all down!

San	ji	han	desu. →
three	(o'clock)	thirty	(it) is

Can you change

FRESH SALSA

to

FIRE SALE

in just SIX moves?!

Hint: one move = remove, add, or move a letter

1. Add I 2. Remove S 3. Remove H 4. Remove S 5. Remove A 6. Add E

Source: Blake Allen and Patrick Littell

HOW DOES AUTOCORRECT WORK?

Edit distance is used in computer science to tell how different two pieces of text are.

Each time you remove, add, or move a letter, it adds one to the edit distance.

In the problem you just solved, for example, the edit distance between FRESH SALSA and FIRE SALE was 6.

Autocorrect works by identifying misspelled words (by choosing words that don't match the list in its dictionary), then changes them to a similar word – that is, a word with a small edit distance.

Spell check works this way too, though it gives you a list of options to choose from in the order of smallest to largest edit distance.

Autocorrect and spell check need to know more than just edit distance. They also need to know which spelling errors are most likely and which typos are most likely.

CAN YOU GUESS THE LANGUAGE?

HINT #1: It's an indigenous language of the United States.

HINT #2: Its writing system looks like this:

ᏣᎳᎩ ᎦᏬᏂᎯᏍᏗ

HINT #3: The name of the language, in the language, is Tsalagi.

Cherokee

Source: www.omniglot.com

Professor Word has invented a machine to read the newspaper aloud, but it isn't working right!

This ad it just read doesn't make any sense…

> "Do you love docks? We11 tick-tock, time is running out Sor dock wor1d's spring sale!
>
> We have watches, grandSather docks, and so much more!
>
> Whether you are a dock co11ector or just buying one Sor Sun, stop by dock wor1d today!"

Can you figure out which THREE MISTAKES the machine is making?

1. f → S
2. l → 1
3. cl → d

HOW DOES TEXT RECOGNITION WORK?

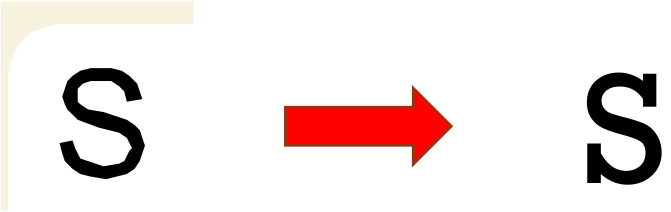
ALSO KNOWN AS OPTICAL CHARACTER RECOGNITION OR OCR

When you type into a computer, the letters are stored as codes in the computer's memory.

When you scan a page into the computer, it creates a picture of the words stored as pixels.

Text recognition was invented to change pixels in into codes for letters.

The computer 'looks' at each black squiggle in the image and tries to match it to one of the letters in its list.



Because a computer can't actually read, though, sometimes it makes mistakes! Unusual fonts and very small letters are a challenge for the program.

The puzzle you just solved is based on real mistakes made by a text recognition program!

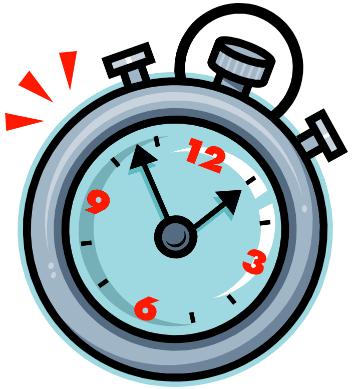
Here are some sentences in Estonian, a language spoken in Estonia (a country in Northeastern Europe).

Kell on üks.	It is one o'clock.
Kell on kaks.	It is two o'clock.
Kell on veerand kaks.	It is quarter past one. ('quarter toward two o'clock')
Kell on pool kaks.	It is half past one. ('half before two o'clock')
Kell on kolmveerand kaks.	It is quarter to two. ('three quarters toward two o'clock')

How would you say "It's quarter past four" in Estonian?

Hint: 'five' = viis

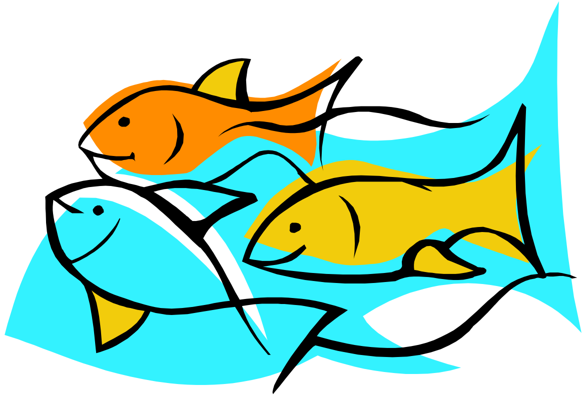
Kell on veerand viis.



Source: Babette Verhoeven-Newsome

Here are some sentences in Hindi, a language spoken in India.

Char matchliyan hain.	There are four fish.
Char ladkiyan hain.	There are four girls.
Che matchliyan hain.	There are six fish.



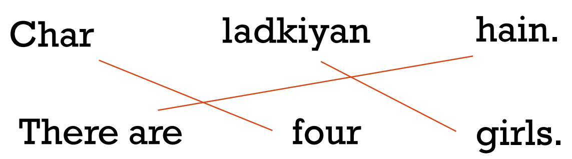
How would you say "There are six girls" in Hindi?

Che ladkiyan hain.

HOW DOES COMPUTER TRANSLATION WORK?

The way you solved this puzzle is the same way a computer translator works!

Translation systems are "trained" to notice words and their translations, and the differences in the order of words.



Char ladkiyan hain.

There are four girls.

We "train" them by programming them count many millions of word correspondences like the ones in this example. After they count, they compute probabilities.

After training, the computer makes a translation dictionary like this:

che = six
ladkiyan = girls
hain = are

The real translation dictionary is full of errors, but it contains a probability for each translation.

When you put in a sentence, the system translates the words using its dictionary and then puts them in the right order. This way, it can translate sentences it has never seen before, without a human having to write them all down!

There are six girls. →

hain che ladkiyan

Fill in the blanks so that both phrases make sense!

BASEBALL BAT CAVE

POOL PARTY HAT

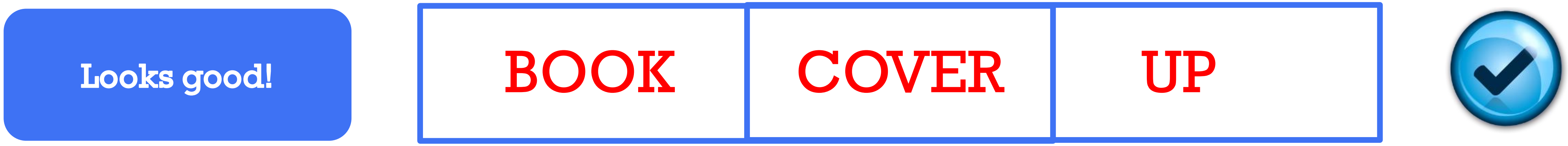
BOOK COVER UP

HOW DO COMPUTERS 'GUESS' WORDS?

FOR SPEECH RECOGNITION AND COMPUTER TRANSLATION

Sometimes, a computer can't be sure what word we just said, or which version of a translation is the right one. But how do you teach a machine to guess, and guess well?!

Computers guess words using probabilities. The program 'looks' at the words before and after the mystery word, then creates a list of all possible words that could go in that blank. Then, it chooses the one that is the most probable with the word before it, and the word after it:



Looks good!

BOOK	COVER	UP

✓

Unfortunately, even when each word in the chain makes sense with the word before it, the whole chain can end up being gibberish! Linguists and computer scientists are working to 'teach' computers better ways to guess words, and in the meantime we can enjoy all the funny mistakes computers make.

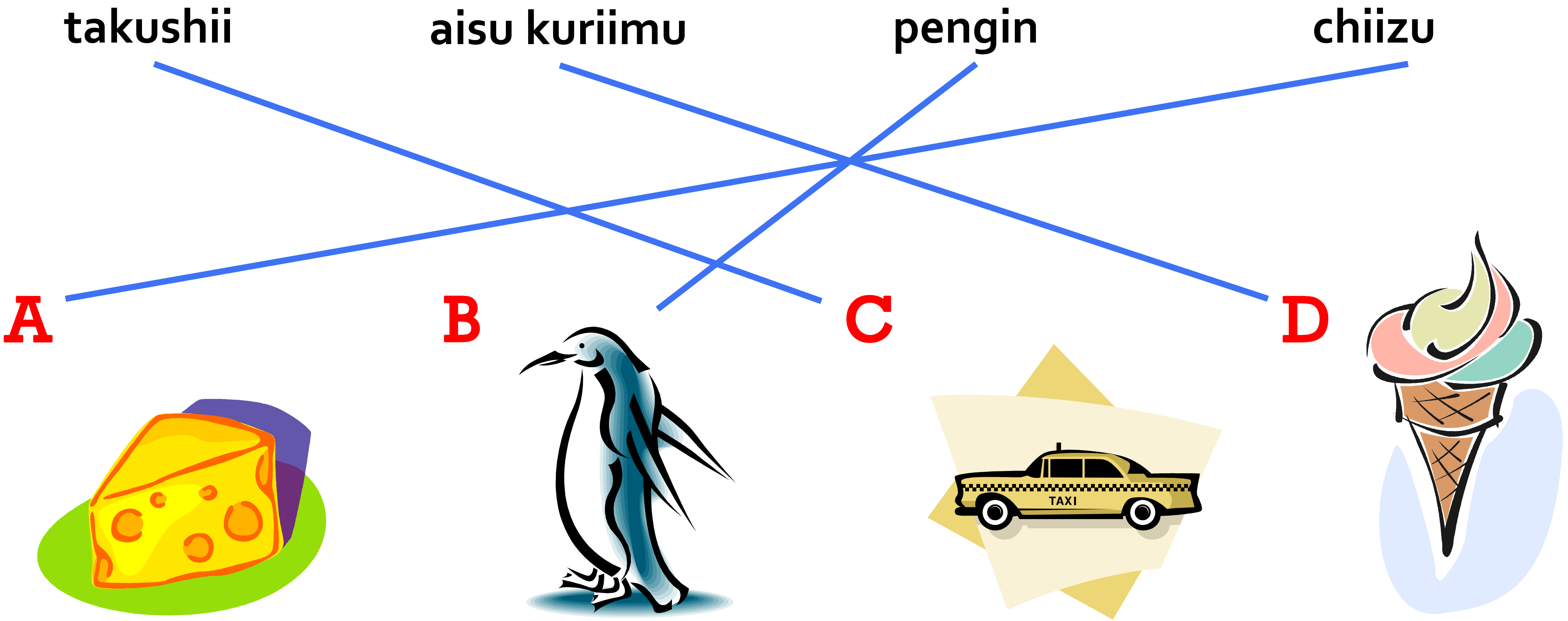
HOW DO COMPUTERS GUESS WORDS?

FOR SPEECH RECOGNITION AND COMPUTER TRANSLATION

- Let’s look at an example from a state-of-the-art computer translation system.
  - Supplied by Austin Matthews.
- The red parts of the sentence are each ok on their own, but together they make a grammatical error:
  - the curriculum will be more emphasis on " real life " problems.
    - Compare to: the curriculum will be more advanced.
  - the curriculum will be more emphasis on " real life " problems.
    - Compare to: the solution will be more emphasis on real life problems.

Some words in Japanese, called loanwords, are “borrowed” from English.

Can you match each loanword to its picture?



takushii

aisu kuriimu

pengin

chiizu

A

B

C

D

WHY DO LANGUAGES SOUND DIFFERENT?

Even when a word is the same in English and another language, it might sound very different!

By the time you're six months old, you can already tell the difference between all the sounds of your native language. Not all languages have the same sounds, though!

In this puzzle, you learned that Japanese puts a vowel after every consonant (except n). So 'taxi' becomes 'takushii'!

Just like the sound 'x' is difficult for Japanese speakers to pronounce, some sounds might seem unusual to you:

- Nepali speakers use four different kinds of 't'!
- Xhosa has three different clicking sounds that are used as letters!
- Some languages in Central Asia can start a word with four consonants!

Each of these newspaper headlines can have two different meanings!

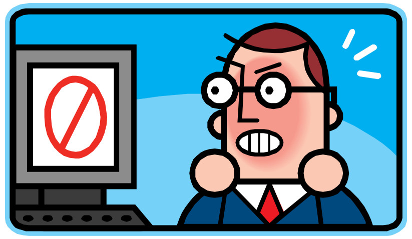
Can you figure out what they are?

A. POP STAR CHASED BY FAN

B. THIEF CAUGHT BY BRIDGE

C. BIG WIN STARTS SEASON

WHY IS IT SO HARD FOR COMPUTERS TO UNDERSTAND US?



Even though it's funny to imagine the hidden meaning of these sentences, you can probably guess which meaning is correct.

We make thousands of those guesses every day -- we don't always say exactly what we mean, but luckily everyone's brain can fill in the gaps.

Unfortunately, the guesses that are so easy for us are very hard for a computer!
In this example, who is smart and who has computers?

[smart students] and [teachers with computers]

[[smart students] and teachers] with computers

smart [students and [teachers with computers]]

smart [students and teachers] with computers

This made up example is simple compared to what computers really encounter in Wikipedia, social media, email, and on-line newspapers. Ambiguity of this sort is *combinatoric* and can fill up a computer's memory quickly.

People, on the other hand, read over most ambiguity without noticing. Why doesn't it fill up your memory? That's a good research question!

Japanese uses a system of letters known as kanji. Each kanji has a specific meaning and pronunciation(s), and kanji can be combined to make new meanings.

食べ = “to eat”

物 = “thing”

What do you think 食べ物 means?

“food”

Japanese uses a system of letters known as kanji. Each kanji has a specific meaning and pronunciation(s), and kanji can be combined to make new meanings.

外 = “outside”

国 = “country”

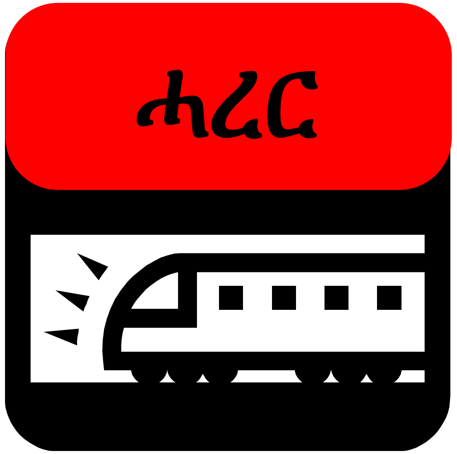
What do you think 外国 means?

“foreign country”

You are in Addis Abeba in Ethiopia. You need to get to Adama. Can you figure out which way to go just by looking at these Amharic signs?

Hint: Amharic is written using the Ge’ez script. In those letters Addis Abeba is spelled አዲስ አበባ !

1



Harar

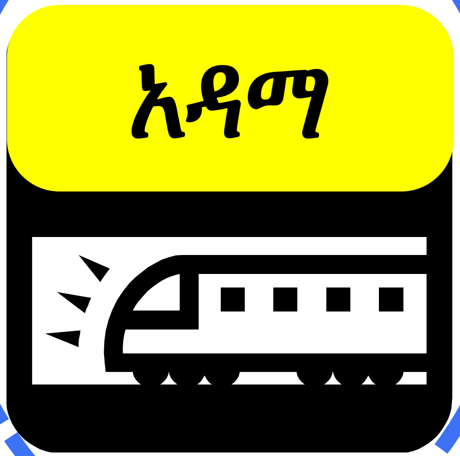
2

Dire Dawa

3

Asaita

4



HOW DO OTHER WRITING SYSTEMS WORK?

In these puzzles you saw that some languages aren't written the same way English is.

Linguists divide writing systems into several different categories:

Alphabets use sets of letters to write consonants and vowels: ABCDE

English

Abjads use sets of letters to write only consonants: קומה יהוה ויפצו

Hebrew

Syllabaries use one letter to represent each syllable: すべてのにんげん

Japanese

Bengali

Abugidas combine consonant symbols with vowel symbols to make each letter: সমস্ত মানুষ স্বাধীনভাবে

Semanto-phonetic systems have many letters, each with its own sound and meaning: 漢字

Chinese

Compared to some of these examples, the 26 letters we use to write English i ally small number!

Source: www.omniglot.com

Can you change

BOB'S RAFTS

to

BARB'S CRAFTS

in just FOUR moves?!

Hint: one move = remove, add, or move a letter

1. Remove O 2. Add A 3. Add R 4. Add C

Source: Blake Allen and Patrick Littell

HOW DOES AUTOCORRECT WORK?

Edit distance is used in computer science to tell how different two pieces of text are.

Each time you remove, add, or move a letter, it adds one to the edit distance.

In the problem you just solved, for example, the edit distance between BOB'S RAFTS and BARB'S CRAFTS was 4.

Autocorrect works by identifying misspelled words (by choosing words that don't match the list in its dictionary), then changes them to a similar word – that is, a word with a small edit distance.

Spell check works this way too, though it gives you a list of options to choose from in the order of smallest to largest edit distance.

Autocorrect and spell check need to know more than just edit distance. They also need to know which spelling errors are most likely and which typos are most likely.

Professor Word has invented a machine to read the newspaper aloud, but it isn’t working right! This story it just read doesn’t make any sense…

> “New bond releoses loue song
>
> A local bond hos just releosed lts flrst slngle, a loue song wrltten by the gultorlst. The song ls olreody #2 on the chorts. When osked how they feel obout thelr newfound success, the whole bond wos speechless! Thelr full album comes out loter thls month.”

Can you figure out which THREE MISTAKES the machine is making?

1. a → o   2. i → l   3. v → u

HOW DOES TEXT RECOGNITION WORK?

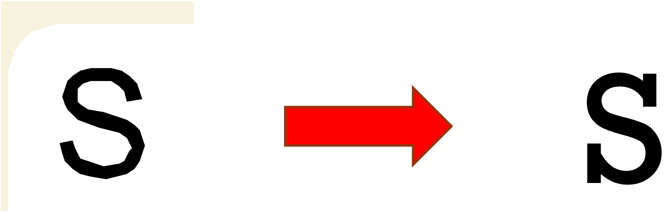
ALSO KNOWN AS OPTICAL CHARACTER RECOGNITION OR OCR

When you type into a computer, the letters are stored as codes in the computer's memory.

When you scan a page into the computer, it creates a picture of the words stored as pixels.

Text recognition was invented to change pixels in into codes for letters.

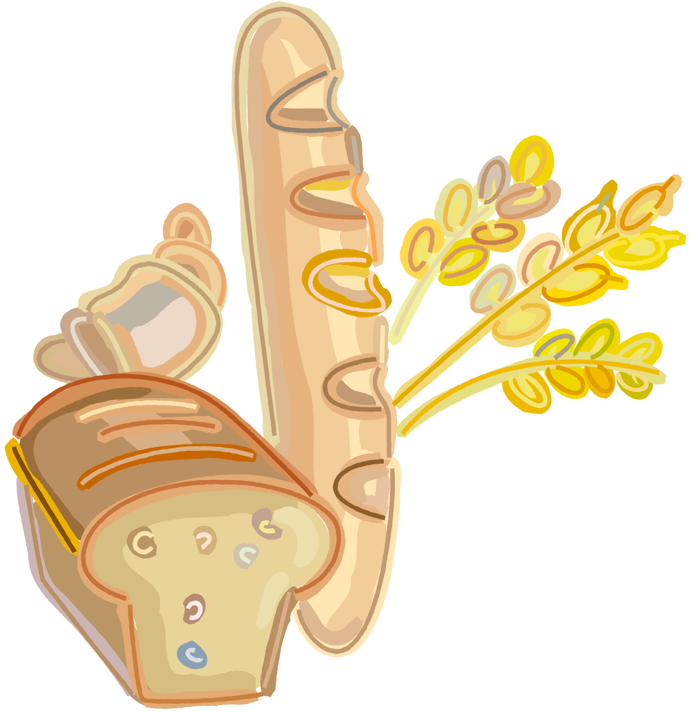
The computer 'looks' at each black squiggle in the image and tries to match it to one of the letters in its list.



Because a computer can't actually read, though, sometimes it makes mistakes! Unusual fonts and very small letters are a challenge for the program.

The puzzle you just solved is based on real mistakes made by a text recognition program!

Here are some sentences in Hindi, a language spoken in India.



Chai dijie.	Tea, please.
Roti dijie.	Bread, please.
Chawal dijie.	Rice, please.

How would you say "Water, please" in Hindi?

Hint: 'water' = pani

Pani dijie.

Here are some sentences in Japanese. Just like in America, Japanese schools have many different levels.

Emi wa chuugakusei desu.	Emi is a middle school student.
Ken wa daigakusei desu.	Ken is a high school student.
Sayuri wa shougakusei desu.	Sayuri is an elementary school student.

If the Japanese word for "I" is watashi, can you say what kind of student you are?

Watashi wa ___ desu.

Here are some sentences in Quechua, a language spoken in Mexico.

Pay asin.	He laughed (very recently).
Pay asiran.	He laughed (a while ago).
Pay asisqa.	He laughed (a very long time ago).
Pay tiyan.	She sat down (very recently).
Pay tiyaran.	She sat down (a while ago).

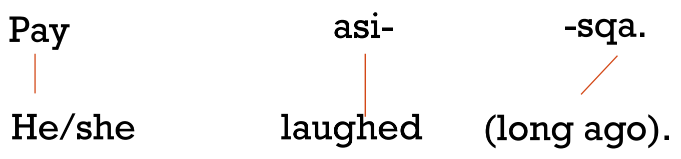
How would you say "She sat down (a very long time ago)" in Quechua?

Pay tiyasqa.

HOW DOES COMPUTER TRANSLATION WORK?

The way you solved this puzzle is the same way a computer translator works!

Translation systems are "trained" to notice words and their translations, and the differences in the order of words.



Pay asi- -sqa.

He/she laughed (long ago).

We "train" them by programming them count many millions of word correspondences like the ones in this example. After they count, they compute probabilities.

After training, the computer makes a translation dictionary like this:

pay = he/she
tiya- = to sit
-sqa = (long ago)

The real translation dictionary is full of errors, but it contains a probability for each translation.

When you put in a sentence, the system translates the words using its dictionary and then puts them in the right order. This way, it can translate sentences it has never seen before, without a human having to write them all down!

She sat (a very long time ago). →

pay tiya sqa

DID YOU KNOW?

There are over 7,000 languages spoken around the world.

Over 382 of those languages are spoken in the United States!

Greek

Vietnamese

Navajo

Persian

Mon-Khmer

French Creole

Hebrew

Spanish

Yiddish

Tagalog

Gujarati

Laotian

Arabic

Armenian

Polish

Thai

Hmong

Source: www.census.gov

Fill in the blanks so that both phrases make sense!

CREDIT CARD GAME

ICE CREAM CHEESE

COUCH POTATO CHIP

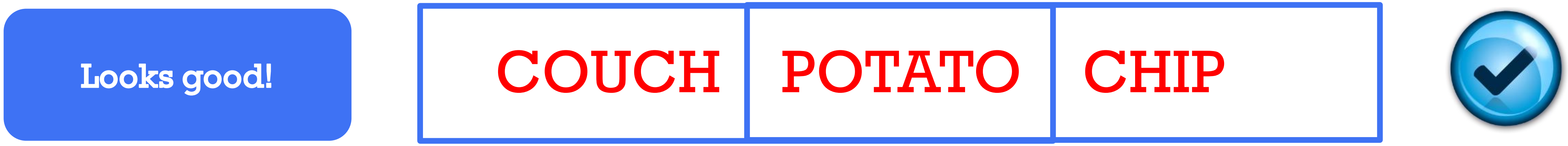
Source: Patrick Littell

HOW DO COMPUTERS 'GUESS' WORDS?

FOR SPEECH RECOGNITION AND COMPUTER TRANSLATION

Sometimes, a computer can't be sure what word we just said, or which version of a translation is the right one. But how do you teach a machine to guess, and guess well?!

Computers guess words using probabilities. The program 'looks' at the words before and after the mystery word, then creates a list of all possible words that could go in that blank. Then, it chooses the one that is the most probable with the word before it, and the word after it:



Looks good!

COUCH POTATO CHIP ✓

Unfortunately, even when each word in the chain makes sense with the word before it, the whole chain can end up being gibberish! Linguists and computer scientists are working to 'teach' computers better ways to guess words, and in the meantime we can enjoy all the funny mistakes computers make.

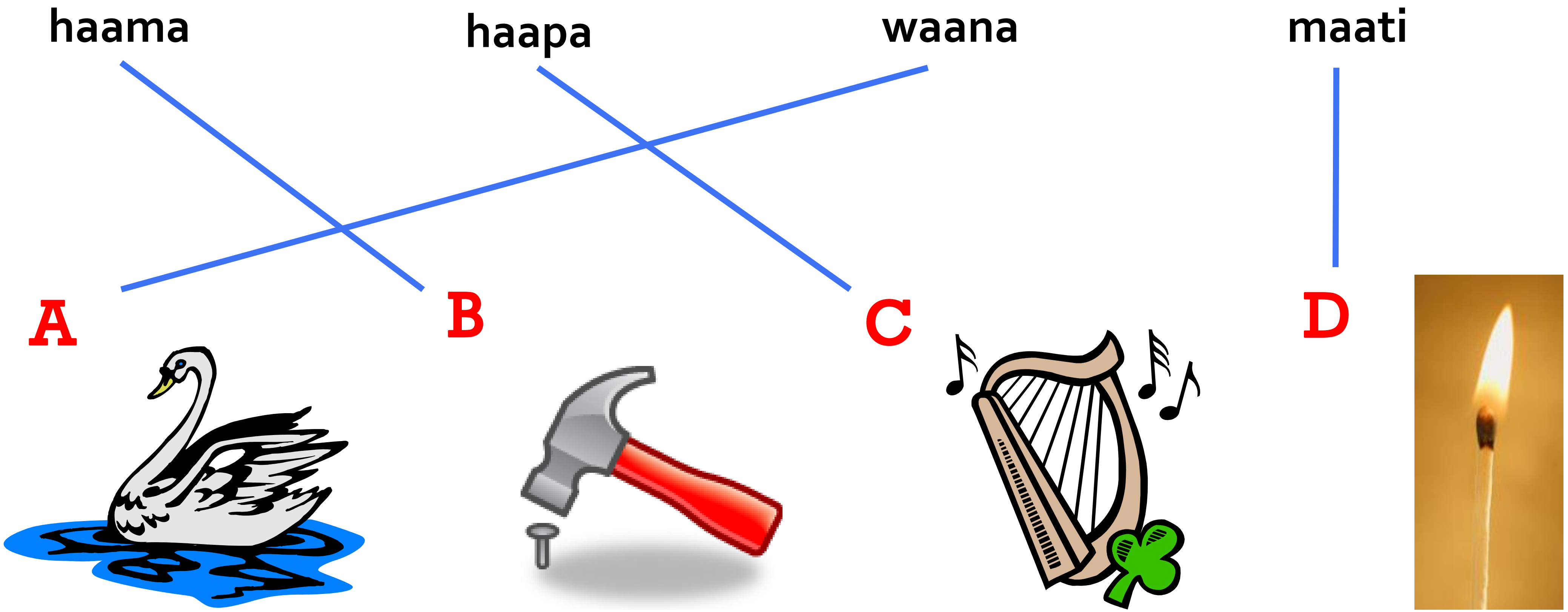
HOW DO COMPUTERS GUESS WORDS?

FOR SPEECH RECOGNITION AND COMPUTER TRANSLATION

- Let’s look at an example from a state-of-the-art computer translation system.
  - Supplied by Austin Matthews.
- The red parts of the sentence are each ok on their own, but together they make a grammatical error:
  - the curriculum will be more emphasis on " real life " problems.
    - Compare to: the curriculum will be more advanced.
  - the curriculum will be more emphasis on " real life " problems.
    - Compare to: the solution will be more emphasis on real life problems.

Maori is a language spoken by the aboriginal (native) people of New Zealand. Some words in Maori, called loanwords, are "borrowed" from English.

Can you match each loanword to its picture?



haama

haapa

waana

maati

A

B

C

D

Source: Patrick Littell

WHY DO LANGUAGES SOUND DIFFERENT?

Even when a word is the same in English and another language, it might sound very different!

By the time you're six months old, you can already tell the difference between all the sounds of your native language. Not all languages have the same sounds, though!

In this puzzle, you learned Maori speakers don't pronounce the letter 's'. So 'swan' becomes 'waana'!

Just like the sound 's' is difficult for Maori speakers to pronounce, some sounds might seem unusual to you:

- Nepali speakers use four different kinds of 't'!
- Xhosa has three different clicking sounds that are used as letters!
- Some languages in Central Asia can start a word with four consonants!

Here are some sentences in Japanese.

Ritsu wa hana ga suki desu.	Ritsu likes flowers.
Chihiro wa hana ga suki jyanai.	Chihiro doesn’t like flowers.
Asako wa ame ga suki desu.	Asako likes candy.

Can you figure out what this Japanese sentence means?

Mizuho wa ame ga suki jyanai.

Mizuho doesn’t like candy.

Here are some sentences in Chol, a language spoken in Mexico.

Mi kocel.	I enter.
Mi ?yocel.	He enters.
Mi kubin.	I listen (to it).
Mi ?yubin.	He listens (to it).

Can you fill in the blank?

Hint: The symbol "?" stands for a glottal stop – the sound you make in the middle of "uh-oh"!

Here are some sentences in Spanish.

El perro duerme.	The dog sleeps.
El perro come.	The dog eats.
El gato duerme.	The cat sleeps.

Can you figure out what this Spanish sentence means?

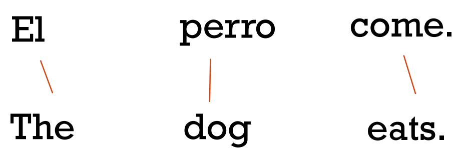
El gato come.

The cat eats.

HOW DOES COMPUTER TRANSLATION WORK?

The way you solved this puzzle is the same way a computer translator works!

Translation systems are "trained" to notice words and their translations, and the differences in the order of words.



We "train" them by programming them count many millions of word correspondences like the ones in this example. After they count, they compute probabilities.

After training, the computer makes a translation dictionary like this:

el = the
gato = cat
come = eats

The real translation dictionary is full of errors, but it contains a probability for each translation.

When you put in a sentence, the system translates the words using its dictionary and then puts them in the right order. This way, it can translate sentences it has never seen before, without a human having to write them all down!

El gato come. →

the cat eats

CAN YOU GUESS THE LANGUAGE?

HINT #1: Its writing system is called Mkhedruli and looks like this:

დამწერლობა

HINT #2: It's spoken in Georgia, Armenia, Azerbaijan, Iran, and other countries in the South Caucasus.

HINT #3: It shares the name of the main country where it's spoken.

Georgian

Source: www.omniglot.com

Each of these newspaper headlines can have two different meanings!

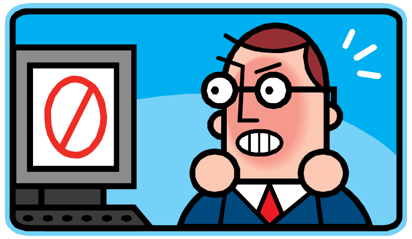
Can you figure out what they are?

A. REWARD OFFERED FOR LOST CAT

B. NEW MALL OPENS DOORS

C. CONVICT BEGINS SENTENCE

WHY IS IT SO HARD FOR COMPUTERS TO UNDERSTAND US?



Even though it’s funny to imagine the hidden meaning of these sentences, you can probably guess which meaning is correct.

We make thousands of those guesses every day -- we don’t always say exactly what we mean, but luckily everyone’s brain can fill in the gaps.

Unfortunately, the guesses that are so easy for us are very hard for a computer!
In this example, who is smart and who has computers?

[smart students] and [teachers with computers]

[[smart students] and teachers] with computers

smart [students and [teachers with computers]]

smart [students and teachers] with computers

This made up example is simple compared to what computers really encounter in Wikipedia, social media, email, and on-line newspapers.   Ambiguity of this sort is *combinatoric* and can fill up a computer’s memory quickly.

People, on the other hand, read over most ambiguity without noticing. Why doesn’t it fill up your memory? That’s a good research question!

Japanese uses a system of letters known as kanji. Each kanji has a specific meaning and pronunciation(s), and kanji can be combined to make new meanings.

今 = "now"

日 = "day"

What do you think 今日 means?

"today"

Japanese uses a system of letters known as kanji. Each kanji has a specific meaning and pronunciation(s), and kanji can be combined to make new meanings.

中 = "middle"

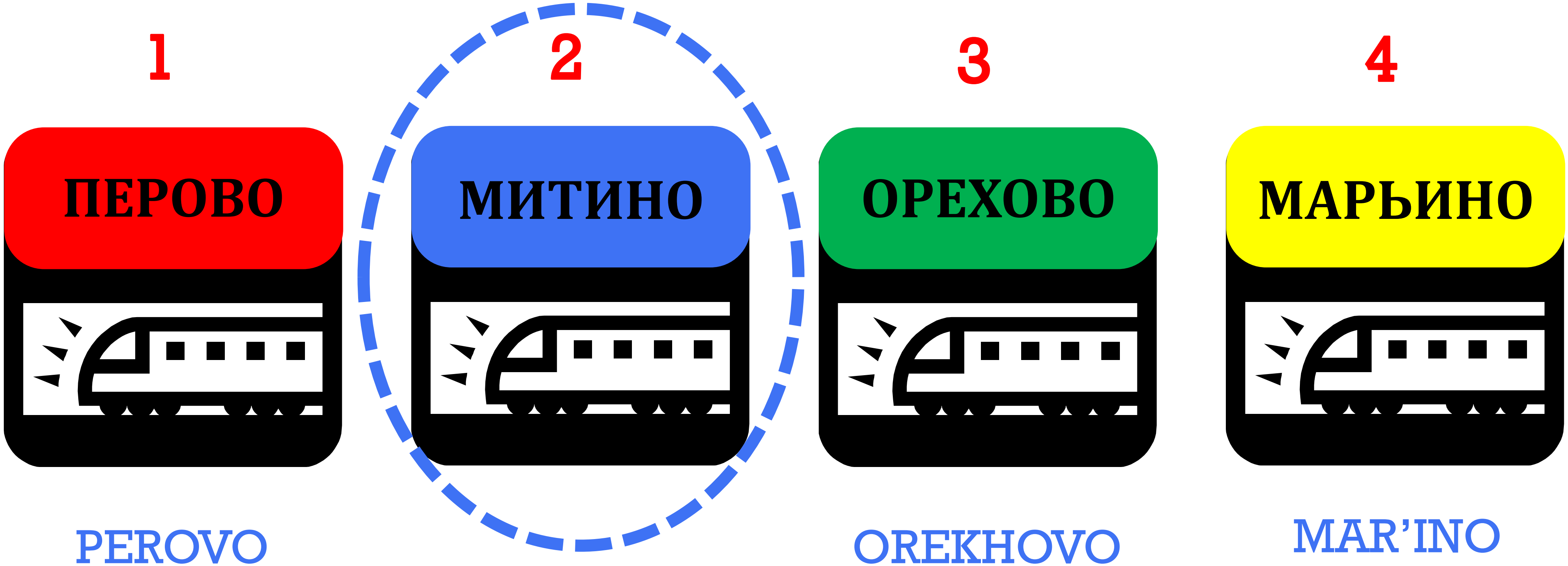
学 = "school, learning"

What do you think 中学 means?

"middle school"

You are in ANNINO Station in Moscow, Russia. You need to get off at the "MITINO" stop. Can you figure out which train to board just by looking at these Russian signs?

Hint: The sign for ANNINO is **АННИНО** in the Russian alphabet!



1 ПЕРОВО PEROVO

2 МИТИНО

3 ОРЕХОВО OREKHOVO

4 МАРЬИНО MAR'INO

HOW DO OTHER WRITING SYSTEMS WORK?

In these puzzles you saw that some languages aren't written the same way English is.

Linguists divide writing systems into several different categories:

Alphabets use sets of letters to write consonants and vowels: ABCDE

Abjads use sets of letters to write only consonants: קומה יהוה ויפצו

Syllabaries use one letter to represent each syllable: すべてのにんげん

Abugidas combine consonant symbols with vowel symbols to make each letter: সমস্ত মানুষ স্বাধীনভাবে

Semanto-phonetic systems have many letters, each with its own sound and meaning: 漢字

English

Hebrew

Japanese

Bengali

Chinese

Compared to some of these examples, the 26 letters we use to write English i ally small number!

Source: www.omniglot.com

HOW DO PUZZLES PREPARE YOU FOR COMPUTING?

- Pattern recognition
- Multi-step reasoning
- Thinking in terms of instructions or procedures
- Thinking of problems as procedures with inputs and outputs
- Breaking complex tasks into simpler tasks

WHY DO COMPUTER SCIENTISTS NEED TO KNOW ABOUT LANGUAGE DIVERSITY?

- Less than half of the Web is in English.
- Computer programs that work for English might not work for other languages if they are not carefully designed.
- Although English is useful in the global economy, local languages preserve identity, cultural heritage, and a legacy of knowledge.
- Even cultures with low literacy have computational needs:
  - They use oral micro-blogs that they access via cell phone.
  - These micro-blogs help them with health care, agricultural information, and weather alerts.
- When you are an executive in a high tech company, will you know how to meet the world's needs?

Careers

Humanitarian

Industry

Government

Academic

Education

Careers

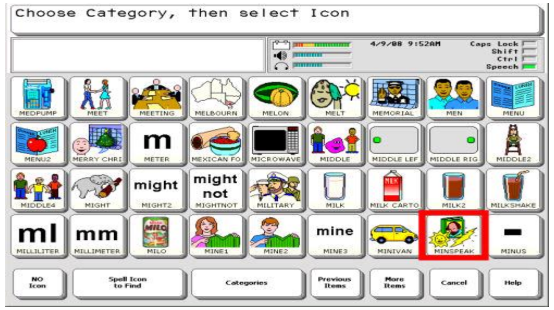
Humanitarian

Machine translation for disaster relief and humanitarian aid.

Translate between aid workers and victims of disease or natural disaster.

Technologies such as spelling checkers to help revitalize endangered languages

Assistive technologies for people with disabilities



Choose Category, then select Icon

Careers

Industry

Facebook
Twitter
Google
Yahoo
Reuters
General Motors
Microsoft
Amazon

Search engines

Natural language voice interfaces
Talking to machines

Summarization
because there is more information than people can attend to

Sentiment detection
Did people like the product or movie?

Machine Translation
Translate from one language to another

Careers

Machine Translation

Speech recognition

Summarization and information extraction

Detection of sentiment and deception

National Security:

There is more information than human analysts can attend to.

Government

Careers

Computer Assisted Language Learning

Automatically detect errors

Automated grading of essays

Educational Testing Service

Analysis of educational dialogue

The way you interact affects the way you learn

Education

Careers

Work at a university

Train the next generation

Do research on unsolved problems in Natural Language Processing

Academic