Hello. I'm Toshiki Maekawa.

I'm a member of Chandrasekhar Project.

Chandrasekhar Project is one of the science clubs in my school.

I'm doing research on natural language processing and understanding.

First, please look at this sentence. 'HOW WILL THE WEATHER BE?'

OK? If you are human, you are able to understand what this sentence means.

But this is a difficult problem for computers.

Because this sentence is just a string of words, so it has no meaning for computers.

These years, computers have been developing.

And getting great results even ordinary people know.

For example, self-driving cars or .

How about the field of 'natural language'? Do you know AI assistants Siri or Google Assistant?

If you've ever used it, you know that they give up understanding meanings of complicated questions.

Then I thought I could do something.

This is why I began to do a research on making computers abstract meanings of sentences.

OK, now I'll explain how it works.

First, ambiguous grammar in natural languege makes it difficult for comuters to abstract the meaning of sentences.

Natural languages used by humans have grammer.

So does programming languages for computers.

But there is an absolute difference. That is, as I said, ambiguousness.

Understanding ambiguous languages is not easy for even humans. But in most cases, it doesn't matter.

It has not been clarified why humans can use ambiguous languages.

But the existence of grammar in natural language means that computers should be able to deal with them.

Indeed, these years, the accuracy of machine translation services has been improving.

If you've not used them, won't you give them a try? You would feel they are more useful than you thought.

I trained a computer with sentences in natural langauge and the lists of meanings in formal language.

But maybe you are not sure what the formal language is.

So I'll explain what it is simply.

Look at the slide.

It may be like instructions group rather than formal language.

For instance, 'weather n' means the weather n days later.

There are more.

In this example, the same color part has the same meaning.

The trained computer in this research will generate this formal language.

Then, how can we make computers be able to abstract the meanings of natural languages?

These years, AI has been developing thanks to the technology called "Deep Learning".

Before explaining Deep Learning, let me explain Neural Network.

Neural Network is a technology that can approach any arbitrary functions by imitating the neural circuits of our brain.

Neural Network is an aggregate of perceptrons.

In the figure of a simple perceptron, X1, X2 and X3 are called 'input',

W1, W2 and W3 are called 'weight',

and Y is called 'output'.

The sum total of multiplied each input and weight is the output.

And correcting the error between the output and the expected output.

This is how the Neural Network works.

Especially in Multilayer Neural Network, its internal structure is devided into 3 layers.

If you have good intuition, maybe you can figure out the 'Deep' in 'Deep Learning' means that there're multiple hidden layers.

Multiple hidden layers need enourmous amount of calculation.

And usually make it so hard to train computers.

So deep Learning was less likely to be used.

But the improvements of computers has made Deep Learning a greatly useful technology.

After that, AI has been developing.

This time, I used Recurrent Neural Network, one of the Deep Learning technologies, in this research.

Recurrent Neural Network can deal with time-series data.

And I especially used Seq-to-Seq Model, one of the Recurrent Neural Network, to train a computer.

It's often used for machine translation these years.

I trained the computer with the model with sentences and the formal language I just explained.

And the computer will be able to output the proper formal language from even unknown input.

But this model has a problem.

It is so bad at training of long inputs.

In order to make up for the problem, Attention Mechanism is used.

This is to train each correspondence between input and output.

Herewith, the accuracy of learning improves.

Most of input sentences contain what most of you ask to AI assistant.

For example, about specific contents like "How is the weather?" or "tell me my plans for tomorrow."

But I couldn't prepare so many sentences.

Because I had to make meaning lists of sentences manually.

For that reason, I used 40 sentences to train the computer.

You probably thought that it's too few. You're right, it may not be enough.

Preparing so many sentences that fits my purpose is difficult.

In addition, it takes a long time to train the computer.

These problems are important issues for the next research.

OK, I'll explain the result.

I gave the trained computer new 20 sentences and

counted up how many correct outputs the computer could output.

Then 13 outputs of 20 were correct.

It's 65%.

I think it is not bad in the dataset as small as 40 sentences.

I show some samples given to the trained computer to recognize.

As they show, sentences contained in dataset such as 'how will the weather be' or ‘what time is it’ are , of course, recognized correctly.

Some sentences not contained in dataset are also recognized correctly.

They show that the computer is trained nicely.

On the other hand, there are some samples not recognized correctly.

'Can you tell me how the weather is' and so on.

But 'tell me how the weather will be' is recognized correctly.

What I want to say is, 'can you tell me how the weather is' and 'tell me how the weather will be' have 'tell me how the weather' in common.

And the common part has the most part of meaning of the sentence.

That means the computer couldn't see which part is concentrated on the meaning of sentence.

If I could prepare more sample sentences, probably they make the accuracy of understanding of meanings better.

And I think there's another problem.

Humans, especially children, come to be able to learn languages without bothering to tell the meanings.

However, I bothered to prepare lists of meanings to train the computer this time.

Common sense tells that it's distinctly inefficient.

Although there is no help unless having been clarified the mechanism of humans' languages,

in order for computers to really understand languages, a drastic change of technology is required.

I want to continue this research to find such a new great technology.