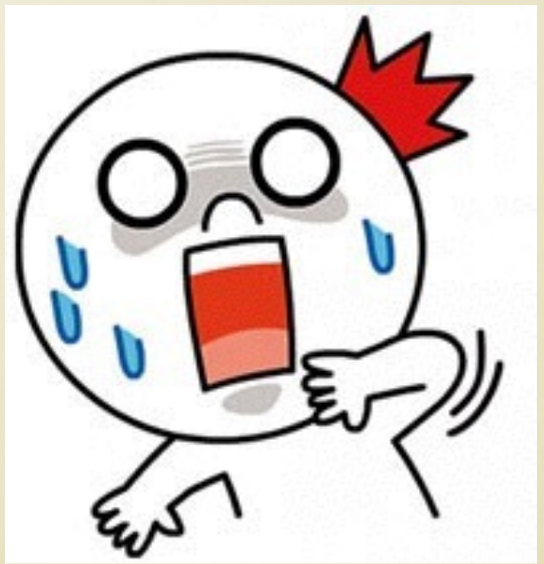
Automatic estimation of meaning ambiguity of emoticons based on linguistic expressibility

Naoto Ishii, Fumito Masui, Michal Ptaszynski
Department of Computer Science,
Kitami Institute of Technology, Japan

Introduction

Online communication has been very popular

(´・ω・`)



Introduction

(*^◯^*)

ヽ(´ー｀)ノ

Sometimes the meaning of emoticons is ambiguous

Introduction

(° A °;)

Emoticons alone sometimes do not convey the whole nuance

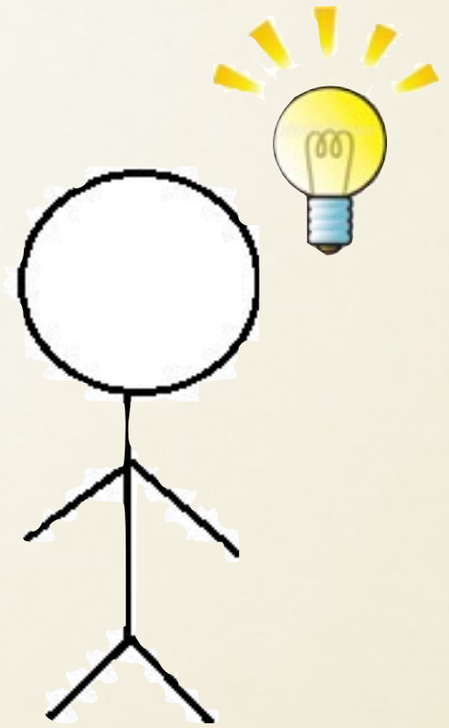
→Cause misunderstandings

Introduction

(゜A゜;)

ゴクリ

Gokuri(gulp)



Emoticons are often used together with onomatopoeia

The relevance between emoticons and onomatopoeia has not yet been made clear sufficiently

Introduction

- **Investigate the relationship between emoticons and onomatopoeia**
- **Propose a method for quantifying whether an emoticon is ambiguous or easy to understand**

Preliminary survey

Questionnaire design

Investigate the relationship between emoticons and onomatopoeia

Question 1

What onomatopoeia do you recall when looking at the following emoticons?
You can answer "I don't know" if you do not recall any.

(1)(つ´∀`)つ

(2) (・∀・)

(3) (＝３＝)

Question 2

Look at the next onomatopoeia. What emoticons do
You can answer "I do not know" if you do no
Also, if you do not know emoticons, you c

2ちゃんねる顔文字　【ｶﾞｰﾝ】

ｶﾞ――(;ﾟ Д ﾟ)――ﾝ!!

ｶﾞ――(ﾟ Д ﾟ;)――――ﾝ!!!!!

ｶﾞ――Σ Σ(ﾟ Д ﾟ;)――ﾝ

Σ(ﾟ Д ﾟ lll)ｶﾞｰﾝ

kaomoji.uunyan.com

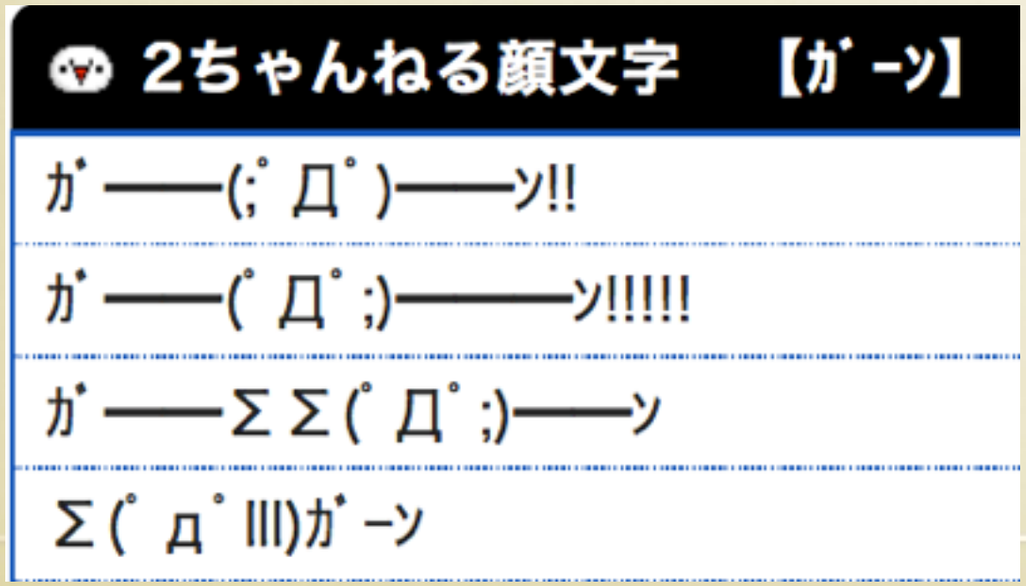
Questionnaire design

Investigate the relationship between emoticons and onomatopoeia

Question 2

Look at the next onomatopoeia. What emoticons do you recall?
You can answer “I do not know” if you do not recall any.
Also, if you do not know emoticons, you can express them with pictures.

(1)*gaaan*(shock)
(2)*doyon*(feel down)
(3)*shakin*(feel refreshed)



kaomoji.uunyan.com

Questionnaire implementation

Target of questionnaire
85 people(70 male, 15 female)

Number of questions in the questionnaire: 20
question 1: 10, question 2: 10

Results

Question 1 (9) (;゜Д゜) (10) Σ(゜д゜lll)

Often assigned words like “gaaan”(shock)

Question 2 (1) gaaan

34 cases out of 51 votes emoticons containing the following eye-mouth-eye triplet were used “ ゜ д ゜ ”

゜д ゜→ Best represented by onomatopoeia “gaaan”

Results

Question 1 (3) (= 3 =)

- Representing **anger or irritation** “Boooooo”
- Representing feelings of **affection** “kiss”

The opinions were not consistent

⇩

Emoticons have various meanings

Results

Emoticons → Onomatopoeia

“maji!” (“no way!”)

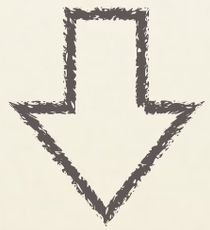
“oteage” (“I give up”)

Included expressions that were not onomatopoeia

Suggests ambiguity of nuances of emoticons

Discussion

- **Emoticons have various meanings**
- **Emoticons are ambiguous expressions**



Emoticons and onomatopoeia do not correspond ideally one to one

Quantification of linguistic expressibility

Quantification of linguistic expressibility

✦What is **linguistic expressibility**?

- Whether one can express concepts of something with words
- What extent can be expressed in words

Hypothesis

Linguistic expressibility is high

⇩

Meaning ambiguity is low

Percentage of “I don’t know”

Rank	Emoticon	Percentage of “I don’t know”(%)
1	(つ´∀`)つ	40.00
1	(;´ρ`)	40.00
3	ヽ(´ー`)ノ	37.65
	:	
8	(;゜Д゜)	17.65
9	＼(*^0^*)／	16.47
10	Σ(゜д゜lll)	5.88
	Average	27.53

The ease of understanding of emoticons depends on the symbols used

Perc f "I don't know"

Eyes and mouth using the symbols ° Д°

→Easy to recall meanings

I don't know"(%)
40.00
40.00
37.65
17.65
16.47
5.88
27.53

Eyes using the symbols ´ `

→Difficult to recall meanings

The ease of understanding of emoticons depends on the symbols used

Perc f "I don't know"

Eyes and mouth using the symbols ° Д°

→Easy to recall meanings

I don't know"(%)
40.00
40.00

→Each emoticon was divided into symbols

→Each symbol in the emoticon was assigned the ratio of "I don't know"

Eyes using the symbols ´ `

→Difficult to recall meanings

5.88
27.53

The ease of understanding of emoticons depends on the symbols used

Automatic estimation of meaning ambiguity of emoticons

Method 1

Emoticons used in the preliminary survey

(゜∀゜)	32.94
f(^^;)	32.94
(・∀・)	30.59
(=з=)	21.18
(;゜Д゜)	17.65

Ambiguity score
Percentage of "I don't know"

Percentage of "I don't know" for each emoticon is used as the ambiguity score of the emoticon

Method 2

Eyes		Mouth		Others	
= =	21.18	д	11.77	;	30.20
・ ・	30.59	з	21.18	*	16.47
^ ^	24.71	∀	34.51	lll	5.88
´ `	39.22	o	16.47	Σ	5.88
° °	18.82	ρ	40.00	f	32.94

Calculate ambiguity for each part

↓

The average of all parts is the ambiguity score of the detected emoticons

Method 3

First, matching whole emoticons

↓

When not detected, matching based on separate characters

Criteria for ambiguity

Example

(・∀・) 30.59

(;゜Д゜) 17.65

(^ д ^ ;)22.23

Average of estimates

(30.59+17.65+22.23)/3=23.49

(・∀・) 30.59 Ambiguity is **high**

(;゜Д゜) 17.65 Ambiguity is **low**

(^ д ^ ;) 22.23 Ambiguity is **low**

Experiment

Gold standard data collection

- **Three methods were executed for the emoticon database (10,137) used in previous research**
  - Method 1：3
  - Method 2 and 3：**86** ⟵ **Gold standard data**

Methods 2 and 3 could be superior

Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R. and Araki, K. : CAO : A Fully Automatic Emoticon Analysis System, In Proc. of The 24th AAAI Conference on Artificial Intelligence (AAAI-10), pp.1026-1032.

Questionnaire inquiring about ambiguity of emoticons

Do you understand the meaning of the following emoticons?
If you do, circle “yes” and write meaning next to the emoticon.
If you do not understand, circle “no”.

(1)d(゜∀ ゜;)　　yes　no

(2)ヽ(°Д°)ノ　　yes　no

(3) (#`・д・)/　yes　no

(4) (艸°Д°＊)　　yes　no

(5) ヽ(°∀°)ノ　　yes　no

Questionnaire inquiring about ambiguity of emoticons

Number of “don't understand” for each emoticon

Number of answers	Number of emoticon	
3	15(17%)	
2	29(34%)	Ambiguity is **high**
1	23(27%)	Ambiguity is **low**
0	19(22%)	

Results and discussion

Performance comparison for each method

$$\text{Precision} = \frac{\text{Emoticons that were consistent with questionnaire results}}{\text{Successfully matching of emoticons}}$$

$$\text{Recall} = \frac{\text{Emoticons that were consistent with questionnaire results}}{\text{Emoticons used in questionnaire}}$$

$$\text{F score} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Performance comparison for each method

Method 1

Precision	Recall	F score
0.33	0.01	0.02

Method 2 and 3

Precision	Recall	F score
0.44	0.44	0.44

Dividing emoticons by symbols

→Can help estimate the ambiguity of emoticons more accurately

Discussion

✦Comparison of Method 1 and Methods 2 and 3

- Method 2 and 3 successfully extracted more emoticons
- Better results were obtained for method 2 and 3 than for method 1 for all scores: Precision, Recall, and F score

Estimation method of based on separate characters could be superior

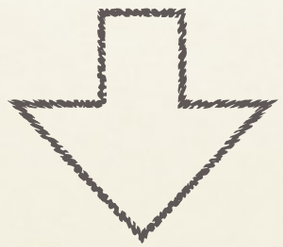
Approximately half of emoticons are ambiguous

Emoticons are highly ambiguous expressions

Conclusions

Conclusions

- Investigate the relationship between emoticons and onomatopoeia
- Propose a method for quantifying whether an emoticon is ambiguous or easy to understand



✦**Emoticons are highly ambiguous expressions**
✦**Estimation method based on separate characters could be superior**

Future task: verify the ambiguity of emoticons when used in wider contexts (sentences, paragraphs, etc.)

Supplementary material

How we choice emoticons and onomatopoeia from “emoticon channel”

- Frequently used emoticons and onomatopoeia
- Little used emoticons and onomatopoeia

Each five was used

Classification of representation of answer items

Difficult to express onomatopoeia

f proportion of expressions not poeic of each emoticon

	Emoticon	Percentage of expression not onomatopoeia(%)
1	(゜∀゜)	61.40
2	(つ´∀`)つ	58.82
2	(;´ρ`)	58.82
	:	
8	f(^^;)	29.82
9	Σ(゜д゜lll)	25.00
0	(＝з＝)	16.42
	Average	43.27

Easy to express onomatopoeia

Standard deviation

We were trying calculate of standard deviation, but many emoticons were neutral

High ambiguity emoticons : 12
Low ambiguity emoticons : 15

About 70% emoticons are neutral

↓

We classified the ambiguity of emoticons to high or low

Interpretation differences by symbols

(^ д ^ *)

musu

tere

runrun

ho

happy ?

relieve ?

annoy ? shame ?

The meaning recalled by the symbol of interest changes

What is Precision, Recall, and F score?

✦Precision

- Number of cases for which correct ambiguity was estimated within detected emoticon

✦Recall

- The number of cases for which correct ambiguity was estimated for in all emoticons included in the gold standard data

✦F score

- The harmonic mean of Precision and Recall

Comparison of Method 2 and 3

- **The emoticon with different results in Method 2 and Method 3**

- ♪(・∀・)つ旦←**There was no difference**
- ヽ(゜∀゜)メ
- ヽ(゜∀゜)ノ

Comparison of Method 2 and 3

✦In **Questionnaire**

- ヽ(゜∀゜)メ
- ヽ(゜∀゜)ノ

✦In **Method 2**

- The ambiguity of both emoticons was **low**

✦In **Method 3**

- The ambiguity of both emoticons was **high**

Re-estimation

Method 2

Precision	Recall	F score
0.59	0.31	0.41

Method 3

Precision	Recall	F score
0.57	0.30	0.39

Method 2 seems to be superior to Method 3

Re-estimation

ヽ(゜∀゜)ノ

✦**Questionnaire：Ambiguity is low**

- Method 2：Ambiguity is **low**
- Method 3：Ambiguity is **high**

It is necessary to perform additional experiments in the future

Final destination of this research

Create to translation system of emoticons - onomatopoeia

How create?

To connect one emoticons and 2 or 3 onomatopoeia, and vice versa

ex.1) (= з =) → “Boooooo” or “kiss” or “musu”

ex.2) Gaaan → “(;゜Д゜)” or “Σ(゜д゜lll)”

Connect those with high voting rate of the questionnaire