

# Annotating Affective Information on 5.5 Billion Word Corpus of Japanese Blogs

Michal Ptaszynski <sup>1</sup>, Pawel Dybala <sup>4</sup>,  
Rafal Rzepka <sup>2</sup>, Kenji Araki <sup>2</sup>, Yoshio Momouchi <sup>3</sup>

1) JSPS Research Fellow / Hokkai-Gakuen University, High-Tech Research Center

2) Hokkaido University, Graduate School of Information Science and Technology

3) Hokkai-Gakuen University, Department of Electronics and Information Engineering

4) Otaru University of Commerce

# Annotating Affective Information on 5.5 Billion Word Corpus of Japanese Blogs

**Michal Ptaszynski**<sup>1</sup>, Pawel Dybala<sup>4</sup>,  
Rafal Rzepka<sup>2</sup>, Kenji Araki<sup>2</sup>, Yoshio Momouchi<sup>3</sup>

1) JSPS Research Fellow / Hokkai-Gakuen University, High-Tech Research Center

2) Hokkaido University, Graduate School of Information Science and Technology

3) Hokkai-Gakuen University, Department of Electronics and Information Engineering

4) Otaru University of Commerce

# Presentation Outline

- Introduction
- Emotion Blog Corpora
- Affective Information Annotation
- Evaluation of Affective Annotations
- Statistics of Affective Annotations
- Conclusions and Future Work

# Introduction

- Recently there has been much research done on Sentiment Analysis and Affect Analysis
  - Sentiment Analysis (SA): Positive/Negative
  - Affect Analysis (AA): Emotion types

# Introduction

- Recently there has been much research done on Sentiment Analysis and Affect Analysis
    - Sentiment Analysis (SA): Positive/Negative
    - Affect Analysis (AA): Emotion types
- “John loves his red shiny Porsche boxster.”



# Introduction

- Recently there has been much research done on Sentiment Analysis and Affect Analysis
    - Sentiment Analysis (SA): Positive/Negative
    - Affect Analysis (AA): Emotion types
- “John loves his red shiny Porsche boxster.”

SA: attitude=Positive, opinion  
object=Porsche boxster (red)

AA: emotion=love, emotion  
object=Porsche boxster (red)



# Introduction

- Blogs have been recognized as a good source of information on attitudes and emotions.

# Introduction

- Blogs have been recognized as a good source of information on attitudes and emotions.

- [1] Quan, C. and Ren, F. 2010. "A blog emotion corpus for emotional expression analysis in Chinese", *Computer Speech & Language*, Vol. 24, Issue 4, pp. 726-749.
- [2] Hashimoto, C., Kurohashi, S., Kawahara, D., Shinzato, K. and Nagata, M. 2011. "Construction of a Blog Corpus with Syntactic, Anaphoric, and Sentiment Annotations" [in Japanese], *Journal of Natural Language Processing*, Vol. 18, No. 2, pp. 175-201.
- [3] Aman, S. and Szpakowicz, S. 2007. "Identifying Expressions of Emotion in Text", *LNAI 4629*, pp. 196-205.
- [4] 阿部修也, 江口萌, 隅田飛鳥, 大崎梓, 乾健太郎. みんなの経験: ブログから抽出したイベントおよびセンチメントのDB化. 言語処理学会第15回年次大会発表論文集, pp.296-299, March 2009.
- [5] Fujimura, S., Fujimura, K. Okuda, H., "Blogosonomy: Autotagging Any Text Using Bloggers' Knowledge", IEEE/WIC/ACM International Conference on Web Intelligence, pp. 205 - 212, 2007,
- [6] G. Leshed and J. Kaye, "Understanding how bloggers feel: recognizing affect in blog posts", CHI'06 extended abstracts on Human factors in computing systems, pp. 1019-1024, 2006.
- [7] J. Li and F. Ren, "Emotion Recognition from Blog Articles", IEEE NLP-KE 2008, pp. 355-362, Beijing, Oct. 2008.
- [8] R. Mihalcea and H. Liu. "A corpus-based approach to finding happiness", In Proceedings of Computational approaches for analysis of weblogs, AAAI Spring Symposium, 2006.
- [9] Dipankar Das, Sivaji Bandyopadhyay, "Labeling Emotion in Bengali Blog Corpus ? A Fine Grained Tagging at Sentence Level", Proceedings of the 8th Workshop on Asian Language Resources, pages 47-55, 2010.



# Introduction

- It would be good to have a big blog corpus annotated with emotions/attitudes.

# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora

# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora

[10] Changqin Quan, Fuji Ren, A blog emotion corpus for emotionalexpression analysis in Chinese, Computer Speech & Language, Volume 24, Issue 4, October **2010**, Pages 726-749.

[11] Janyce Wiebe, Theresa Wilson, and Claire Cardie (**2005**). Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, volume 39, issue 2-3, pp.165-210.

[12] Chikara Hashimoto, Sadao Kurohashi, Daisuke Kawahara, Keiji Shinzato and Masaaki Nagata, "Construction of a Blog Corpus with Syntactic, Anaphoric, and Sentiment Annotations" [in Japanese], Journal of Natural Language Processing, Vol 18, No. 2, pp. 175-201, **2011**.

[13] J. Minato, D. B. Bracewell, F. Ren and S. Kuroiwa, "Statistical Analysis of a Japanese Emotion Corpus for Natural Language Processing", LNCS 4114, **2006**.

[14] Aman, S. and Szpakowicz, S. (**2007**). Identifying Expressions of Emotion in Text. In Proceedings of the 10th International Conference on Text, Speech, and Dialogue (TSD-2007), Plzen, Czech Republic, Lecture Notes in Computer Science (LNCS), Springer-Verlag.

[15] Dipankar Das, Sivaji Bandyopadhyay, "Labeling Emotion in Bengali Blog Corpus - A Fine Grained Tagging at Sentence Level", Proceedings of the 8th Workshop on Asian Language Resources, pages 47-55, **2010**.

[16] Kazuyuki Matsumoto, Yusuke Konishi, Hidemichi Sayama, Fuji Ren, "Analysis of Wakamono Kotoba Emotion Corpus and Its Application in Emotion Estimation", International Journal of Advanced Intelligence, Vol. 3, No. 1, pp.1-24, March, **2011**.

[17] Gilad Mishne, Experiments with Mood Classification in Blog Posts. In: Style2005 the 1st Workshop on Stylistic Analysis of Text for Information Access, at SIGIR 2005, August **2005**.

# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora

	Number of sentences		number of documents
[10]	12,724		500
[11]	10,657		535
[12]	4,186		249
[13]	1,191		(separate sentences)
[14]	5,205		173
[15]	12,149		123
[16]	4,773		(separate sentences)
[17]	?		815,494

# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora

Number of sentences | number of documents

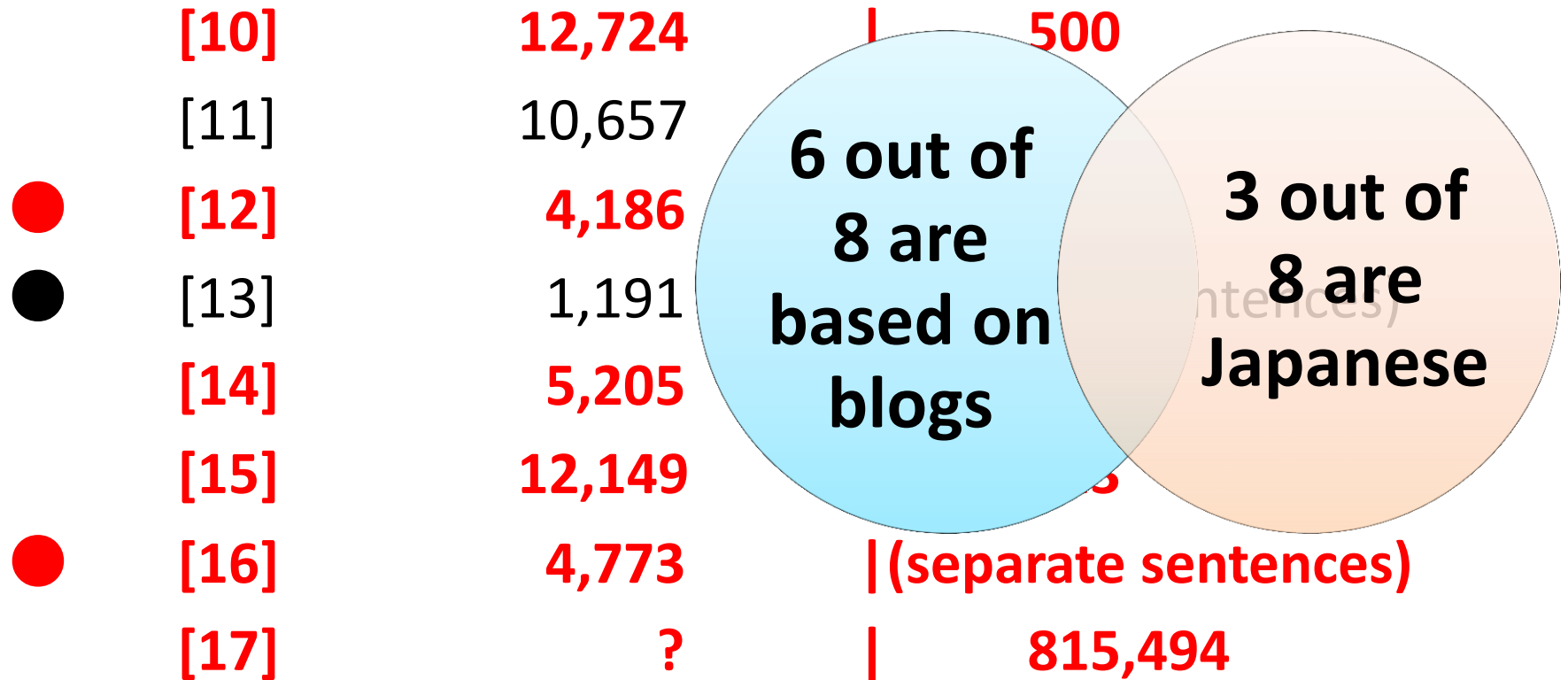
[10]	12,724		500
[11]	10,657		
[12]	4,186		
[13]	1,191		(sentences)
[14]	5,205		
[15]	12,149		
[16]	4,773		(separate sentences)
[17]	?		815,494

6 out of  
8 are  
based on  
blogs

# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora

Number of sentences | number of documents



# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora
  - **Emotion blog corpora are rather small**
  - **Japanese emotion blog corpora are the smallest**

# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora  
Amount of  
annotations



# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora

Amount of annotations

	emotion classes	annotated affective information					syntactic annotations
		emotive expressions	emotive/non-emot.	valence/activation	emotion intensity	emotion objects	
[10]	8	○	○	○/×	○	○	T,POS;
[11]	0	○	○	○/×	○	○	T,POS;
● [12]	0	○	×	○/×	×	○	T,POS,L, DP,NER;
● [13]	8	○	○	×/×	×	×	POS;
[14]	6	○	○	×/×	○	×	×
[15]	6	○	×	×/×	○	×	×
● [16]	9	○	×	×/×	×	×	×
[17]	123	×	×	×/×	×	×	×

# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora

Some are focused on syntax/morphology annotations

	emotion classes	annotated affective information					syntactic annotations
		emotive expressions	emotive/non-emot.	valence/activation	emotion intensity	emotion objects	
[10]	8	○	○	○/x	○	○	T,POS;
[11]	0	○	○	○/x	○	○	T,POS;
● [12]	0	○	×	○/x	×	○	T,POS,L, DP,NER;
● [13]	8	○	○	x/x	x	x	POS;
[14]	6	○	○	x/x	○	x	x
[15]	6	○	x	x/x	○	x	x
● [16]	9	○	x				
[17]	123	x	x	x/x			

Some are focused on emotion annotations

# Emotion Blog Corpora

- We need an emotion corpus that is:
  1. Large
  2. Annotated with syntax/morphology
  3. Annotated with emotions

# Emotion Blog Corpora

- YACIS corpus
  1. 354 mil. Sentences in 13 mil. Documents.
  2. Tokenized, lemmatized, annotated with POS, dependency structure, named entities.

# Emotion Blog Corpora

- YACIS is:
  1. Large
  2. Annotated with syntax/morphology
  3. Annotated with emotions

# Affective Information Annotation

- Tools
  - ML-Ask: Affect Analysis system for text-based utterances
  - CAO: Emoticon Analysis system

# Affective Information Annotation

- ML-Ask
  - Distinguish emotive sentences from non-emotive
  - Annotate emotive expressions in emotive sentences
  - Generalize emotive expressions into:
    - Emotion classes
    - Valence/activation

[1] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "Affecting Corpora: Experiments with Automatic Affect Annotation System - A Case Study of the 2channel Forum -" The Conference of the Pacific Association for Computational Linguistics (PACLING-09), September 1-4, 2009, Hokkaido University, Sapporo, Japan, pp. 223-228.

[2] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, "A System for Affect Analysis of Utterances in Japanese Supported with Web Mining", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Special Issue on Kansei Retrieval, Vol. 21, No. 2 (April), pp. 30-49 (194-213), 2009.

# Affective Information Annotation

- ML-Ask

**High Accuracy**

- Distinguish emotive sentences from non-emotive

- Annotate emotive expressions in emotive sentences

**High Precision, Low Recall**

- Generalize emotive expressions into:

- Emotion classes
- Valence/activation

[1] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "Affecting Corpora: Experiments with Automatic Affect Annotation System - A Case Study of the 2channel Forum -" The Conference of the Pacific Association for Computational Linguistics (PACLING-09), September 1-4, 2009, Hokkaido University, Sapporo, Japan, pp. 223-228.

[2] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, "A System for Affect Analysis of Utterances in Japanese Supported with Web Mining", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Special Issue on Kansei Retrieval, Vol. 21, No. 2 (April), pp. 30-49 (194-213), 2009.



# Affective Information Annotation

- CAO
  - Detect emoticons in sentence
  - Extract/analyze emoticons
  - Annotate sentences containing emoticons with
    - Emotion classes
    - Valence/activation

[1] Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics", In IEEE Transactions on Affective Computing, vol. 1, no. 1, pp. 46-59, Jan.-June 2010

[2] Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "CAO: A Fully Automatic Emoticon Analysis System", In Proceedings of The Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10), pp. 1026-1032, July 11 - 15, 2010, Atlanta, Georgia, USA

[3] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "Towards Fully Automatic Emoticon Analysis System (^o^)", In Proceedings of The Sixteenth Annual Meeting of The Association for Natural Language Processing (NLP-2010), pp. 583-586, 2010.

# Affective Information Annotation

- CAO

High Accuracy

- Detect emoticons in sentence

- Extract/analyze emoticons

High Accuracy

- Annotate sentences containing emoticons with

- Emotion classes

- Valence/activation \*) **condition: sentence must contain emoticon**

[1] Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics", In IEEE Transactions on Affective Computing, vol. 1, no. 1, pp. 46-59, Jan.-June 2010

[2] Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "CAO: A Fully Automatic Emoticon Analysis System", In Proceedings of The Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10), pp. 1026-1032, July 11 - 15, 2010, Atlanta, Georgia, USA

[3] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "Towards Fully Automatic Emoticon Analysis System (^o^)", In Proceedings of The Sixteenth Annual Meeting of The Association for Natural Language Processing (NLP-2010), pp. 583-586, 2010.

# Affective Information Annotation

- System output

Input sentence: なぜかレディーガガを見ると恐怖感じる(;´艸`)

**ML-Ask output:**

なぜかレディーガガを見ると**恐怖**感じる(;´艸`)

**sentence: emotive**

**emotemes: EMOTICON:(;´艸`)**

**emotions:(1),FEAR:恐怖**

**2D: NEGATIVE, ACTIVE**

**CAO output:**

**Extracted emoticon: (;´艸`)**

**Emoticon segmentation:**

S <sub>1</sub>	BL	S <sub>2</sub>	ELMER	S <sub>3</sub>	BR	S <sub>4</sub>	.
N/A	(		;	'艸'		N/A	)   N/A

**Emotion score**

**Fear (0.02708333)**

Surprize (0.01973684)

Dislike (0.0105364)

Excitement (0.01018174)

Anger (0.00703125)

Sorrow (0.004665203)

Shame (0.004424779)

Joy (0.002962932)

Fondness (0.001851166)

Relief (0)

# Affective Information Annotation

- System output

Input sentence: なぜかレディーガガを見ると恐怖感じる(;´艸`)

**ML-Ask output:**

なぜかレディーガガを見ると**恐怖**感じる(;´艸`)

sentence: emotive

emotemes: **EMOTICON**:(;´艸`)

emotions:(1),**FEAR**:恐怖

**2D: NEGATIVE, ACTIVE**

**CAO output:**

Extracted emoticon: (;´艸`)

Emoticon segmentation:

S <sub>1</sub>	BL	S <sub>2</sub>	ELMER	S <sub>3</sub>	BR	S <sub>4</sub>	.
N/A	(		;	´艸`	N/A	)	N/A

**Emotion score**

**Fear (0.02708333)**

Surprize (0.01973684)

Dislike (0.0105364)

Excitement (0.01018174)

Anger (0.00703125)

Sorrow (0.004665203)

Shame (0.004424779)

Joy (0.002962932)

Fondness (0.001851166)

Relief (0)

**Both systems have been  
evaluated separately, but  
not together  
(CAO supports ML-Ask)**

# Evaluation of Affective Annotations

- Data set
  - Applied earlier in evaluation of CAO in [1]
  - 1000 random sentences from YACIS
  - 418 emotive, 582 non-emotive
  - One part:
    - 42 laypeople annotated emotion classes on sentences

# Evaluation of Affective Annotations

- Results

	emotive/ non-emotive	emotion classes	2D (valence and activation)
<b>ML-Ask</b>	98.8%	73.4%	88.6%
<b>CAO</b>	97.6%	80.2%	94.6%
<b>ML-Ask+CAO</b>	100.0%	89.9%	97.5%

# Evaluation of Affective Annotations

- Results

On blogs many sentences contain emoticons

ML-Ask was better on blogs than on original dataset \*

	emotive/ non-emotive	emotion classes	2D (valence and activation)
<b>ML-Ask</b>	98.8%	73.4%	88.6%
<b>CAO</b>	97.6%	80.2%	94.6%
<b>ML-Ask+CAO</b>	100.0%	89.9%	97.5%

Both systems combined were always better

**On YACIS blog corpus the two systems together have good results.**

\* Proof that dataset influences results.

# Statistics of Affective Annotations



# Statistics of Affective Annotations

- Emotive / Non emotive

---

# of emotive sentences	233,591,502
# of non-emotive sentence	120,408,023
ratio (emotive/non-emotive)	1.94

---

- 2 times more emotive sentences than non-emotive

# Statistics of Affective Annotations

- Emotion classes

# Statistics of Affective Annotations

- Emotion classes
- Three emotion classes were dominant
  - Joy 喜 (+), dislike 厭 (-), fondness 好 (+)
  - Japanese generally express more positive emotions on blogs

in YACIS corpus

emotion class	# of sentences	%	emotion class	# of sentences	%
joy	16,728,452	31%	excitement	2,833,388	5%
dislike	10,806,765	20%	surprize	2,398,535	5%
fondness	9,861,466	19%	gloom	2,144,492	4%
fear	3,308,288	6%	anger	1,140,865	2%
relief	3,104,774	6%	shame	952,188	2%

# Statistics of Affective Annotations

- Emotion classes
- Three emotion classes were dominant
  - Joy 喜 (+), dislike 厭 (-), fondness 好 (+)
  - Japanese generally express more positive emotions on blogs

\* In previous research, on *2channel* two dominating emotions were: dislike (-) and excitement (+/-)

Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "Affecting Corpora: Experiments with Automatic Affect Annotation System - A Case Study of the 2channel Forum -" The Conference of the Pacific Association for Computational Linguistics (PACLING-09), September 1-4, 2009, Hokkaido University, Sapporo, Japan, pp. 223-228.

in VACIS corpus

# of sentences	%
33,388	5%
98,535	5%
44,492	4%
40,865	2%
52,188	2%

# Statistics of Affective Annotations

- Emotion classes
- Three emotion classes were dominant
  - Joy 喜 (+), dislike 厭 (-), fondness 好 (+)
  - Japanese generally express more positive emotions on blogs

\* In previous research, on *2channel* two dominating emotions were: dislike (-) and excitement (+/-)

Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, "Affecting Corpora: Experiments with Automatic Affect Annotation System - A Case Study of the 2channel Forum -" The Conference of the Pacific Association for Computational Linguistics (PACLING-09), September 1-4, 2009, Hokkaido University, Sapporo, Japan, pp. 223-228.

in VACIS corpus

	# of sentences	%
	33,388	5%
	98,535	5%
	44,492	4%
	40,865	2%
	52,188	2%

# Statistics of Affective Annotations

- Emotion classes

In lexicon:

emotion class	number of expressions	emotion class	number of expressions
dislike	532	fondness	197
excitement	269	fear	147
sadness	232	surprise	129
joy	224	relief	106
anger	199	shame	65

**No significant correlation between number of words in lexicon and frequency of emotion class! \***

**\*  $\rho=0.38$**

in YACIS corpus

**Its good because it means lexicon does not influence the results.**

emotion class	# of sentences	%	emotion class	# of sentences	%
joy	16,728,452	31%	excitement	2,833,388	5%
dislike	10,806,765	20%	surprize	2,398,535	5%
fondness	9,861,466	19%	gloom	2,144,492	4%
fear	3,308,288	6%	anger	1,140,865	2%
relief	3,104,774	6%	shame	952,188	2%

# Statistics of Affective Annotations

- Comparison with other corpora

# Statistics of Affective Annotations

- YACIS and KNB
  - KNB:
    - 249 pages, 67,000 words
    - No emotion types, but valence/attitude and opinion-related annotations
    - Compared:  
POSITIVE SENTENCES vs. NEGATIVE SENTENCES  
in YACIS and KNB

**KNB:** Chikara Hashimoto, Sadao Kurohashi, Daisuke Kawahara, Keiji Shinzato and Masaaki Nagata, “Construction of a Blog Corpus with Syntactic, Anaphoric, and Sentiment Annotations” [in Japanese], Journal of Natural Language Processing, Vol 18, No. 2, pp. 175-201, 2011.



# Statistics of Affective Annotations

- YACIS and KNB
- Similar ratio was observed for the two blog corpora: large (YACIS) and small (KNB).

		positive	negative	ratio
KNB*	emotional attitude	317	208	1.52
	opinion	489	289	1.69
	merit	449	264	1.70
	acceptation or rejection	125	41	3.05
	event	43	63	0.68
	sum	1,423	865	1.65
	<hr/>			
YACIS**	only	22,381,992	12,837,728	1.74
	(ML-Ask) only+ mostly	23,753,762	13,605,514	1.75
(ML-Ask-simple)	only	31,071,945	17,496,901	1.78
	only+ mostly	32,752,589	18,442,602	1.78

\* p<.05, \*\* p<.01

**KNB:** Chikara Hashimoto, Sadao Kurohashi, Daisuke Kawahara, Keiji Shinzato and Masaaki Nagata, “Construction of a Blog Corpus with Syntactic, Anaphoric, and Sentiment Annotations” [in Japanese], Journal of Natural Language Processing, Vol 18, No. 2, pp. 175-201, 2011.

# Statistics of Affective Annotations

- YACIS and KNB

**Japanese generally  
express more  
positive emotions  
on blogs.  
(confirmation)**

		positive	negative	ratio	
KNB*	emotional attitude	317	208	1.52	
	opinion	489	289	1.69	
	merit	449	264	1.70	
	acceptation or rejection	125	41	3.05	
	event	43	63	0.68	
	sum	1,423	865	1.65	
	<hr/>				
YACIS** (ML-Ask)	only	22,381,992	12,837,728	1.74	
	only+ mostly	23,753,762	13,605,514	1.75	
	<hr/>				
	(ML-Ask-simple)	only	31,071,945	17,496,901	1.78
only+ mostly		32,752,589	18,442,602	1.78	

\* p<.05, \*\* p<.01

**KNB:** Chikara Hashimoto, Sadao Kurohashi, Daisuke Kawahara, Keiji Shinzato and Masaaki Nagata, “Construction of a Blog Corpus with Syntactic, Anaphoric, and Sentiment Annotations” [in Japanese], Journal of Natural Language Processing, Vol 18, No. 2, pp. 175-201, 2011.

# Statistics of Affective Annotations

- YACIS and Minato et al.
  - Corpus by Minato et al. (2006) [1]
    - Examples (sentences) gathered from Emotion Dictionary [2] and analyzed.
    - Annotations (in general):
      - Emotion classes on sentences
      - POS

[1] Minato, J., Bracewell, D. B., Ren, F. and Kuroiwa, S. 2006. "Statistical Analysis of a Japanese Emotion Corpus for Natural Language Processing", *LNCS 4114*, pp. 924-928.

[2] I. Hiejima. *A short dictionary of feelings and emotions in English and Japanese*, Tokyodo Shuppan, 1995.

[3] Nakamura, A. 1993. *Kanjo hyogen jiten*, Tokyodo Publishing

# Statistics of Affective Annotations

- YACIS and Minato et al.
  - Compared: DISTRIBUTION OF EMOTION CLASSES
  - Between: YACIS, Minato et al. and Nakamura's Emotion Dictionary[3] \*

\* ) dictionary used as affect lexicon in ML-Ask

[1] Minato, J., Bracewell, D. B., Ren, F. and Kuroiwa, S. 2006. "Statistical Analysis of a Japanese Emotion Corpus for Natural Language Processing", *LNCS 4114*, pp. 924-928.

[2] I. Hiejima. *A short dictionary of feelings and emotions in English and Japanese*, Tokyodo Shuppan, 1995.

[3] Nakamura, A. 1993. *Kanjo hyogen jiten*, Tokyodo Publishing

# Statistics of Affective Annotations

- YACIS vs. Minato et al. vs. Nakamura

	Minato et al.	YACIS	Nakamura
dislike	355	14,184,697	532
joy	295	22,100,500	224
fondness	205	13,817,116	197
sorrow	205	2,881,166	232
anger	160	1,564,059	199
fear	145	4,496,250	147
surprise	25	3,108,017	129
	Minato et al. and Nakamura	Minato et al. and YACIS	YACIS and Nakamura
Spearman's $\rho$	0.88	0.63	0.25

[1] Minato, J., Bracewell, D. B., Ren, F. and Kuroiwa, S. 2006. "Statistical Analysis of a Japanese Emotion Corpus for Natural Language Processing", *LNCS 4114*, pp. 924-928.

[2] I. Hiejima. *A short dictionary of feelings and emotions in English and Japanese*, Tokyodo Shuppan, 1995.

[3] Nakamura, A. 1993. *Kanjo hyogen jiten*, Tokyodo Publishing

# Statistics of Affective Annotations

- YACIS vs. Minato et al. vs. Nakamura

1. No significant correlation between: YACIS and Nakamura \*

	Minato et al.	YACIS	Nakamura
dislike	355	14,184,697	532
joy	295	22,100,500	224
fondness	205	13,817,116	197
sorrow	205	2,881,166	232
anger	160	1,564,059	199
fear	145	4,496,250	147
surprise	25	3,108,017	129
	Minato et al. and Nakamura	Minato et al. and YACIS	YACIS and Nakamura
Spearman's $\rho$	0.88	0.63	0.25

\* confirmation of previous calculation good because lexicon does not influence the results

[1] Minato, J., Bracewell, D. B., Ren, F. and Kuroiwa, S. 2006. "Statistical Analysis of a Japanese Emotion Corpus for Natural Language Processing", *LNCS 4114*, pp. 924-928.

[2] I. Hiejima. *A short dictionary of feelings and emotions in English and Japanese*, Tokyodo Shuppan, 1995.

[3] Nakamura, A. 1993. *Kanjo hyogen jiten*, Tokyodo Publishing

# Statistics of Affective Annotations

- YACIS vs. Minato et al. vs. Nakamura

2. Medium correlation between YACIS and Minato (could be some similarities, but nothing 100% sure)

	Minato et al.	YACIS	Nakamura
dislike	355	14,184,697	532
joy	295	22,100,500	224
fondness	205	13,817,116	197
sorrow	205	2,881,166	232
anger	160	1,564,059	199
fear	145	4,496,250	147
surprise	25	2,106,017	129
	Minato et al. and Nakamura	Minato et al. and YACIS	YACIS and Nakamura
Spearman's $\rho$	0.88	0.63	0.25

[1] Minato, J., Bracewell, D. B., Ren, F. and Kuroiwa, S. 2006. "Statistical Analysis of a Japanese Emotion Corpus for Natural Language Processing", LNCS 4114, pp. 924-928.

[2] I. Hiejima. *A short dictionary of feelings and emotions in English and Japanese*, Tokyodo Shuppan, 1995.

[3] Nakamura, A. 1993. *Kanjo hyogen jiten*, Tokyodo Publishing

# Statistics of Affective Annotations

- YACIS vs. Minato et al. vs. Nakamura

3. STRONG correlation between Minato and Nakamura

(both are dictionaries, but differ in: time, media, collecting person's background, approach, assumptions.....)

	Minato et al.	YACIS	Nakamura
dislike	355	14,184,697	532
joy	295	22,100,500	224
fondness	205	13,817,116	197
sorrow	205	2,881,166	232
anger	160	1,564,059	199
fear	145	4,496,250	147
surprise	25	3,108,017	129
	Minato et al. and Nakamura	Minato et al. and YACIS	YACIS and Nakamura
Spearman's $\rho$	0.88	0.63	0.25

[1] Minato, J., Bracewell, D. B., Ren, F. and Kuroiwa, S. 2006. "Statistical Analysis of a Japanese Emotion Corpus for Natural Language Processing", LNCS 4114, pp. 924-928.

[2] I. Hiejima. *A short dictionary of feelings and emotions in English and Japanese*, Tokyodo Shuppan, 1995.

[3] Nakamura, A. 1993. *Kanjo hyogen jiten*, Tokyodo Publishing



# Statistics of Affective Annotations

- What does it all mean?
  - There could be a tendency in Japanese language in general to produce expressions of some emotions (dislike 厭・嫌, excitement 昂, sadness 哀・悲, joy 喜,) than the other (surprise 驚, fear 怖, anger 怒)
  - The number of expressions in language is not equivalent to the frequency of usage
    - Some expressions are used more frequently (relief, fear) than the others (anger sadness)

# Conclusions

- Presented a study on Emotion Corpora with a focus on Emotion Blog Corpora
- Used 2 tools for Affective Information Annotation (ML-Ask and CAO) to Annotate YACIS
  - The tools showed good performance on YACIS
- Presented Statistics of Affective Annotations

# Conclusions

- Presented Statistics of Affective Annotations
  - Japanese express on blogs more positive emotions than negative
  - Japanese language seems to have a tendency to generate more expressions in Japanese for some emotion types than the others
  - However, number of expressions and usage frequency are not related
    - Emotion class with many expressions is sometimes expressed rarely (sadness, anger),
    - Emotion class with few expressions is sometimes expressed more often (relief, fear)

# Future Work

- Online interface!
- More detailed evaluation
- N-gram version for download without limitations
- Applications
  - Affect Analysis
  - Sentiment Analysis
  - Emotion-aware dialog agent development

**Thank you for your attention!**

**Michal Ptaszynski**  
**[ptaszynski@ieee.org](mailto:ptaszynski@ieee.org)**





# Emotion Blog Corpora

- Let's compare Existing Emotion Corpora

corpus name	scale (in sentences / docs)	language	annotated affective information						syntactic annotations
			emotion classes (standard)	emotive expressions	emotive/non-emot.	valence/activation	emotion intensity	emotion objects	
Ren-CECps1.0 [10]	12,724 / 500	Chinese	8 (Yahoo! news annotation standard)	○	○	○/×	○	○	T,POS;
MPQA [11]	10,657 / 535	English	none (no standard)	○	○	○/×	○	○	T,POS;
KNB [12]	4186 / 249	Japanese	none (no standard)	○	×	○/×	×	○	T,POS,L, DP,NER;
Minato et al. [13]	1,191 separate sentences	Japanese	8 (chosen subjectively)	○	○	×/×	×	×	POS;
Aman&Szpakowicz [14]	5205 / 173	English	6 (face recognition)	○	○	×/×	○	×	×
Das&Bandyopadhyay [15]	12,149 / 123	Bengali	6 (face recognition)	○	×	×/×	○	×	×
Wakamono Kotoba [16]	4773 separate sentences	Japanese	9 (6 from face recognition plus 3 added subjectively)	○	×	×/×	×	×	×
Mishne [17]	815,494 blog posts	English	132 (LiveJournal annotation standard)	×	×	×/×	×	×	×