N-gram p ‰ œ p ¸ p ^ [' fi o

„ c S ″ a m Œ ¿ v ° l r ·
„ w H w m \ æ H w ¨
{sainoh,murakami,tokuhisa,ikehara}@ike.tottori-u.ac.jp

1. ″ ¶

@ B | ¤ ¢ ˜ C n Œ A Œ ⁄ ¨ ¡ I ˘ Ł ´ \ » P ° ˘ ˜ | s ⁄ ß @ ‟ ♯ Œ ˜ ¢ Ø D − Œ C fl ¶ ¡ ´ “ œ { Œ \ »” ˘ “ p Œ \ »” ¸ ‰ “ œ p ¸ p ^ [” ° ‹ • Ø K v ‟ Ø D − œ p ¸ p ^ [¡ I ¸ ″ l Ł ¯ » f • Ø ‰ C ° œ p ¸ p ^ [l Ł ¯ ‹ • Ø − ˘ ″ ¢ ¯ Ø D » − ¯ C l Ł ¯ ‹ • Ø ‰ C œ p ¸ p ^ [' fi o • Ø ß @ ‟ K v ˘ ¨ Ø D

] C ¸ R [p X ' œ p ¸ p ^ [' fi o • Ø ß @ ‟ æ ˜ ♯ Œ ˜ « ‰ [1][2] D ' C ' fi o ♯ Œ ‰ œ p ¸ p ^ [″ C P Œ ‹ … ‟ ‰ ' ` ‰ D

» − ¯ { e ¯ ″ C n Œ A Œ ⁄ ¨ ¡ I ˘ Ł ´ \ » ¸ ˘ ‰ œ p ¸ p ^ [' fi o • Ø ß @ æ ˜ • Ø D

2.] Ł @

F [1] ″ C Œ ¯ ¸ t fl ‟ s ƒ Ø P ° j b g ˘ Ł ` C œ { Œ j b g e e Œ ¸ ˜ @ B | ¸ « ' Œ Ø Œ Q ˘ C p Œ j b g e Œ ˘ ¸ x ¯ Ø ¸ m M x C ¸ W Ł ˜ ¢ Ø D

k ″ [2] ″ C œ { Œ \ » ˘ p Œ \ » ¸ t fl Ø ‰ C , C æ ° ˘ Dice W r ŒDice W ß ‟ D Œ ˜ ¢ Ø − ˘ ƒ ‰ D » ˜ C ¸ R [p X ¡ æ o » • Ø C • ♯ P Œ æ ¸ ˜ Dice W ‰ fi Ł ` ¸ W Ł ˜ ¢ Ø D ' C ' fi o ♯ Œ ‰ œ p ¸ p ^ [ ″ C P Œ ‹ … ‟ ‰ ' ` ‰ D

» − ¯ C { e ¯ ″ C n Œ A Œ ⁄ ¨ ¡ I ˘ Ł ´ \ » ¸ ˘ ‰ œ p ¸ p ^ [‹ w C , C ¸ R [p X œ { ¶ ¤ p ¶ » Œ … Œ ' ¡ æ o » • Ø œ { Œ \ » ¤ p Œ \ » o • Ø D » ˜ C … ‟ o ♯ Œ ‰ ¶ W p ¢ ˜ œ { Œ \ » ˘ p Œ \ » ¡ I ¸ t fl Ø ß @ ƒ • D

3. œ p ¸ p ^ [' fi o Ł

œ p ¸ p ^ [' fi o • Ø Ł ¨ ″ ƒ • D

1. œ { Œ \ » o
2. p Œ \ » o
3. œ { Œ \ » ˘ p Œ \ » ¸ t fl » Œ … Œ ´ ¢ ˜ ¨ ″ • Ø D

3.1 œ { Œ \ » o

{ e ¯ ″ C ¸ R [p X ' œ { Œ \ » o • Ø Œ C A ‰ ^ ⁄ N \ » N-gram v ß @ [3] p ¢ Ø D A ‰ ^ ⁄ N \ » N-gram v ß @ ″ C ¡ ¶ ' A − I ¨ ⁄ ° ¶ æ o • Ø ß @ ¯ Ø D } 1 ' ″ uab v ‟ o ♯ Œ Ø D

(1) a b c

(2) a b d

} 1: A ‰ ^ ⁄ N \ » o

A ‰ ^ ⁄ N \ » N-gram v ß @ ″ C o } § ´ ¢ ˜ C ♯ } § ^ C › } § ^ C ª } § ^ 3 ´ ß @ ‟ Ø [4] D

› } § ^ ″ ¡ ‰ ¨ ¢ ¶ æ o • Ø ´ \ « ‟ Æ ¢ ‰ C { e ¯ ″ › } § ^ p ¢ ˜ œ { Œ \ » o • Ø D

3.2 p Œ \ » o

¸ R [p X ' p Œ \ » o • Ø Œ C œ { Œ \ » o • Ø Œ ˜ fl l C A ‰ ^ ⁄ N \ » N-gram v ß @ › } § ^ p ¢ Ø D

3.3 œ { Œ \ » ˘ p Œ \ » ' fi I ¨ ¸ t fl

3.3.1 ¶ Œ v ƒ

{ e ¯ ″ C œ { Œ \ » ¤ p Œ \ » o ″ A ‰

^ ⁄ N \ » *N*-gram v ß @ p ¢ Ø D ' C *N*-gram v ß @ ˝ æ Œ ' fl \ » o • Ø – ˘ ˝ ¯ « ¨ ¢ ‰ C ° X o ‰ œ { Œ \ » ˘ p Œ \ » ¸ t fl Ø K v ‟ Ø D » – ¯ C œ { Œ \ » ¤ p Œ \ » ‟ o † Œ ‰ ¶ W p ¢ ˜ œ { Œ \ » ˘ p Œ \ » ¡ I ¸ t fl Ø D I ˝ C œ { Œ \ » ¶ ¶ ˘ p Œ \ » ¶ ¶ r • Ø D » ˜ C fl ¶ ¶ œ { Œ \ » ˘ p Œ \ » ˝ œ p ¸ p ^ [¯ Ø ´ \ « ‟ ¢ ˘ … Ł C ¶ ‟ Œ v ˜ ¢ Ø ‟ ¢ \ » fl m œ p ¸ p ^ [˘ ˜ ' fi o • Ø [5] D { e ¯ ˝ C ¶ ‟ Œ v ˜ ¢ Ø “ ¶ Œ v ƒ” ˘ Ł ' C fi (1) ¯ Ø D @¶ Œ v ƒ ¶ Œ v ^ o æ c(1) o æ ˘ ˝ C ¸ R [p X ' o † Œ ‰ \ » o æ ¯ Ø D

3.3.2 œ { Œ \ » ˘ p Œ \ » ¸ t fl Ø Ł } 2 ¤ ¢ ˜ C œ p ¸ p ^ [o • Ø ß @ • Ø D } 2 ¯ ˝ C œ { Œ \ » ubc v œ { ¶ ˘ C p Œ \ » uBC v ŒEF v p ¶ ‟ ¸ ¯ Ø (¶ (1)) D ‰ C œ { Œ \ » ubc v œ { ¶ ˘ C p Œ \ » uBC v ŒIJ v p ¶ ‟ ¸ ¯ Ø (¶ (2)) D

(1) : a b c d (1) : A B C D E F

(2) : e b c f (2) : G B C H I J

} 2: ¸ R [p X Æ(1)

} 2 œ { Œ \ » ˘ p Œ \ » ¶ \ 1 C ¶ Œ v ƒ \ 2 ƒ • D

\ 1: œ { Œ \ » ˘ p Œ \ » ¶

œ { Œ \ »	p Œ \ »
ubc v c ¶ (1),(2)	uBC v c ¶ (1),(2) uEF v c ¶ (1) uIJ v c ¶ (2)

\ 2: ¶ Œ v ƒ

œ { Œ \ » p Œ \ » ¶ Œ v ƒ
ubc v uBC v 100% (¶ (1), (2) ⁄ ¿ C(1), (2) ‟ Œ v)
ubc v uEF v 50% (¶ (1), (2) ⁄ ¿ C(1) ‟ Œ v)
ubc v uIJ v 50% (¶ (1), (2) ⁄ ¿ C(2) ‟ Œ v)

» ˜ C ¶ Œ v ƒ ‟ ¢ œ { Œ \ » ubc v ˘ p Œ \ » uBC v œ p ¸ p ^ [˘ • Ø D

4. –

{ e ¯ ˝ C l Ł ¯ œ p ¸ p ^ [‹ • Ø ‰ C ¶ Œ v ƒ C ¶ Œ v ƒ ‟ ¢ œ p ¸ p ^ [˘ ˜ ' fi o • Ø D » ˜ C o ‰ œ p ¸ p ^ [¸ C l Ł ¯] ¿ s ⁄ D

4.1 –

(1) ¸ R [p X ¡ ¸ « [6] ' o ‰ P ¶ ¸ R [p X ˘ ˜ g p • Ø D ¸ R [p X ¶ Ø Æ ¢ T Ø ‰ C10,000 ¶, 50,000 ¶, 80,000 ¶ ¸ ˜ – s ⁄ D Æ \ 3 ƒ • D

\ 3: ¸ R [p X Æ(2)

œ { ¶	p ¶
¢ ˝ ¨ fi ¤ ¯ Ø D	A dog is a faithful animal.
s J \ G ` ‰ D	I bought a Picasso.
˝ X y C Œ ‟ ' Ø D	She understands Spanish.

(2) i u • œ { Œ \ » ¤ p Œ \ » o • Ø Œ C ¡ I ˘ Ł ´ \ » o • Ø ‰ C i † ˘ P Œ u • • Ø ß @ [7] ‟ æ ˜ † Œ ˜ ¢ Ø D { e ¯ ˝ C l X ¨ œ p ¸ p ^ [o • Ø ‰ C ¨ ” 4 ´ Œ ¸ ˜ – s ⁄ D

1. ¶ P ° ˘ ‰ ° Œ
2. … u • ‰ Œ
3. … E fi u • ‰ Œ
4. … E fi E u • ‰ Œ

(3) œ p ¸ p ^ [o œ p ¸ p ^ [¯ Ø ´ \ « ‟ ¢ Ł ƒ · › o • Ø ‰ C ¶ Œ v ‟ 2 ¨ ª ¯ C ' ´ C ¶ Œ v ƒ ‟ 50% ¨ ª œ p ¸ p ^ [˘ ˜ o • Ø D

4.2] ¿ ß @

œ p ¸ p ^ [' fi o ‰ ª C l Ł ¯] ¿ s ¢ C œ p ¸ p ^ [Ł • Ø D] ¿ ˝ _ I æ 50 ´ ¸ ˜ s ¢ C3 ´ ‟ • Ø D @ : fi S ¸ ¯ Ø ˘ » f † Œ Ø @¢ : ¸ ¯ Ø ˘ » f † Œ Ø @~ : ¸ ¯ ˝ ¨ ¢ ˘ » f † Œ Ø] ¿ “ ¢” ˝ C l Ł ¯ C † s ⁄ – ˘ ¯ C œ p ¸ p ^ [‟ ‹ ¯ « Ø ´ \ « ‟ Ø – ˘ ˘ • Ø D » ˜ C † ƒ fi (2) ¯ Ø D @† ƒ “ ” ˘ “ ¢” ^] ¿ ¸ c(2)

5. – °

5.1 † ƒ

{ e ¯ ″ C ¡ ‚ « ' o ‰ P ¶ p ¢ ~ – s ` ‰ D ‰ C – ¯ o ‰ œ p ‚ p ^ [⁄ ¿ C _ I æ 50 ´ ‚ ˜ l Ł ¯] ¿ s ¢ C † ƒ ‰ D

P ¶10,000 ¶ ' o ‰ œ p ‚ p ^ [˘ † ƒ \ 4 ƒ • D ‰ fl l ©50,000 ¶ ˘ † ƒ \ 5 ©80,000 ¶ ˘ † ƒ \ 6 ƒ • D

\ 4: 10,000 ¶ † ƒ

		¢	~	† ƒ	
°	10	37	3	94% (47/50)	112
…	9	36	5	90% (45/50)	133
… E fi	13	29	8	84% (42/50)	113
… E fi E	10	28	12	76% (38/50)	108

\ 5: 50,000 ¶ † ƒ

		¢	~	† ƒ	
°	9	37	4	92% (46/50)	1,317
…	7	35	8	84% (42/50)	1,058
… E fi	5	32	13	74% (37/50)	672
… E fi E	5	29	16	68% (34/50)	613

\ 6: 80,000 ¶ † ƒ

		¢	~	† ƒ	
°	14	34	2	96% (48/50)	2,735
…	7	33	10	80% (40/50)	2,181
… E fi	6	36	8	84% (42/50)	1,296
… E fi E	1	35	14	72% (36/50)	1,198

– ¯ o ‰ œ p ‚ p ^ [″ ©10,000 ¶ ¯ ″ 110 ´ ` ‰ D50,000 ¶ ˘ 80,000 ¶ ¯ ″ ° Œ ‟ ¯ ‰ › C i u • s ⁄ ˘ ‟ › ¨ › ¨ ` ‰ D † ƒ ″ C ° Œ ‟ › C i u • s ⁄ ˘ Æ › ¨ ` ‰ D

S ¯ ″ ©80,000 ¶ ° Œ ¯ o ‰ œ p ‚ p ^ [‟ ‰ › C † ƒ ' ` ‰ D i u • s ` ‰ Œ ″ o ‰ œ p ‚ p ^ [‟ , Ł C † ƒ ″ ‟ ` ‰ D

5.2 œ p ‚ p ^ [Æ

fi S ‚ ¯ Ø ˘ » f ‰] ¿ “ ” Æ \ 7 C ‚ ¯ Ø ˘ » f ‰] ¿ “ ¢” Æ \ 8 C ‚ ¯ ″ ¨ ¢ ˘ » f ‰] ¿ “ ~” Æ \ 9 ƒ • D

\ 7:] ¿ “ ” Æ

	œ { Œ \ »	p Œ \ »
°	• s s ` ‰	went on a trip
…	*N* ″ *N* ‟	*N* is good at *N*
… E fi	*N* ″ *N* *V* Œ ˜ ¢ ‰	the *N* was *V*-ed in *N*
… E fi E	*N* *N* ″ *AdvV* Œ ˜ ¢ Ø	*N*'s *N* is *V*-ed *Adv*

\ 8:] ¿ “ ¢” Æ

	œ { Œ \ »	p Œ \ »
°	– Œ ″	in this case the
…	*N* ″ *N* Ł · ‰	*N* recovered *N*'s *N*
… E fi	*N* ' – *N* *V* ‰	*N* *V*-ed this *N* from *N*
… E fi E	*AdvN* *N* ‟ « ¢	*N* is wrong with *N*'s *N*

\ 9:] ¿ “ ~” Æ

	œ { Œ \ »	p Œ \ »
°	Ł ″ › {	is the duty of
…	*N* ‟ *N*	*N* have put on
… E fi	¤ *N* ‟ *V*	*V* *N* for *N*'s *N*
… E fi E	*N* ‟ *AdvV* °	*N* can *Adv* *V* the *N*

\ 7 \ 8 ' C ¡ I ˘ Ł ´ œ p ‚ p ^ [o ¯ « ‰ – ˘ ‟ ‟ ' ` ‰ D

6. l @

6.1 A ‰ ^ ⁄ N \ » ‚ ‰ œ p ‚ p ^ [

6.1.1 † ƒ

i u • s ` ‰ œ p ‚ p ^ [ ″ C ° Œ o ‰ œ p ‚ p ^ [ ‟ › ¨ › C † ƒ Æ ' ` ‰ D i u • s ⁄ – ˘ ¯ ° æ ‟ , Œ C œ { Œ \ » ˘ p Œ \ » ¡ I ‚ » f ‟ ¢ ` ‰ ‰ ˘ l ƒ Œ Ø D ' C l X ¨ œ p ‚ p ^ [ ‟ K v ˘ ¨ Ø Œ ‟ Ø ‰ C i u • s ` ‰ œ p ‚ p ^ [″ C ¡ ª C œ p ‚ p ^ [‹ • Ø ª ¯ L Ł ¨ w j ¨ Ø – ˘ ‟ œ ¯ « Ø D

6.1.2 œ p ‚ p ^ [‹

{ – ¯ ″ C ‚ ¯ Ø ˘ » f ‰] ¿ “ ¢” œ p ‚ p ^ [‟ ‰ ' ` ‰ D] ¿ “ ¢” œ p ‚ p ^ [″ C l Ł ¯ C † s ⁄ – ˘ ¯ fi S ‚ ¨ œ p ‚ p ^ [‹ • Ø – ˘ ‟ ¯ « Ø D] ¿ “ ¢” œ p ‚ p ^ [l Ł ¯ C † œ p ‚ p ^ [‹ • Ø Œ C l Ł ¯ C † ″ ¢ ¨

Œ ‟ Ø ‟ C œ p ˛ p ^ [‹ ¯ « Ø ˘ l ƒ Œ Ø D

\ 8 ˛ ˜ l Ł ¯ C ł s ` ‰ ° \ 10 ƒ • D ″ ‟ ‟ C ł ¯ Ø D

\ 10: C ł Æ

œ { Œ \ »	p Œ \ »
– Œ ″	in this case (“the”)
N ″ *N* *N* Ł · ‰	*N* recovered *N*’s *N*
N ″*N* ' – *N* *V* ‰	*N V*-ed this *N* from *N*
N ″*AdvN* *N* ‟ « ¢	*N* is *Adv* wrong with *N*’s *N*

6.2 £ U ^ ⁄ N \ » ˛ ‰ œ p ˛ p ^ [

6.2.1 œ p ˛ p ^ [

{ Ł @ ¯ ″ Cbetween ‘ and v ⁄ ¨ £ Œ ‰ Œ Ø £ U ^ ⁄ N \ » ˛ ‰ œ p ˛ p ^ [o ´ \ ¯ Ø D

P ¶80,000 ¶ ° Œ ¯ ″ C œ p ˛ p ^ [3 ´ o ‰ D ° \ 11 ƒ • D ″ › ¨ ¢ ‟ C œ p ˛ p ^ [Ø – ˘ ‟ ¯ « ‰ D

\ 11: £ U ^ ⁄ N \ » °

] ¿	œ { Œ \ »	p Œ \ »
¢	″ A ¢ › ´ ' _ ¯ ‘ ˘ ¨ Ø	differs from ‘ in several ways
¢	† ″ ‘ ° ‟ ″ ¯ Ø	a colonel is ‘ a general
	| ″ · ¶ ‘ ¶ ˜ ¢ Ø	the translation retains the ‘ of the original

6.2.2 œ { Œ \ » ˘ p Œ \ »

£ U ^ ⁄ N \ » ˛ ‰ œ p ˛ p ^ [″ C A ‰ ^ ⁄ N \ » ˛ ‰ Œ C o ‰ œ p ˛ p ^ [‟ › ¨ ' ` ‰ D » – ¯ C P ¶ 80,000 ¶ ° Œ ˛ ˜ C A ‰ ^ ⁄ N \ » ˛ ‰ Œ ˘ £ U ^ ⁄ N \ » ˛ ‰ Œ r s ` ‰ D ° \ 12 ƒ • D

\ 12: A ‰ ^ ⁄ N \ » ˘ £ U ^ ⁄ N \ » r

	œ { Œ \ »	p Œ \ »	
A ‰ ^	13,629	12,691	2,735
£ U ^	1,331	318	3

\ 12 ' C £ U ^ ⁄ N \ » ˛ ˘ ‰ Œ ″ C A ‰ ^ ⁄ N \ » ˛ ˘ ‰ Œ C ˛ R [p X ' ° X o ‰ œ { Œ \ » ˘ p Œ \ » ‟ › ¨ ¢ – ˘ ‟ ‟ ' ` ‰ D ¡ ª ″ C £ U ^ ⁄ N \ » • ‰ C œ { Œ \ » ¤ p Œ \ » ł Ł ƒ · › o • Ø K v ‟ Ø ˘ l ƒ Œ Ø D

7. ¤ Ł

{ e ¯ ″ C l Ł ¯ œ p ˛ p ^ [‹ • Ø ‰ C , C*N*-gram v ß @ p ¢ ˜ C ˛ R [p X ' œ { Œ \ » ¤ p Œ \ » o ‰ D C ¶ Œ v ƒ Ø – ˘ ¯ œ p ˛ p ^ [' fi o ‰ D » ˜ C o ‰ œ p ˛ p ^ [l Ł ¯] ¿ C ł ƒ ‰ D A ‰ ^ ⁄ N \ » ˛ C i u • s ¢ – s ` ‰ ° C l X ¨ ¡ I ˘ Ł ´ œ p ˛ p ^ [ ' fi o • Ø – ˘ ‟ ¯ « ‰ D » Œ … Œ œ p ˛ p ^ [ ˘ ł ƒ ‰ ° Œ80,000 ¶ ° Œ ‟ o ‰ œ p ˛ p ^ [ ‟ ‰ › C ł ƒ ″ 96% ` ‰ D ‰ C S ˜ ł ƒ ‰ ˘ ‰ ˘ – ° C æ8 ‟ l Ł ¯ » f Ł C œ p ˛ p ^ [‹ ‟ ´ \ ¯ Ø – ˘ ‟ ‟ ' ` ‰ D o ‰ œ p ˛ p ^ [ ″ C ˛ ¯ Ø ˘ » f ‰ ] ¿ “ ¢” ‟ ‰ ' ` ‰ D] ¿ “ ¢” œ p ˛ p ^ [″ l Ł ¯ C ł s ⁄ ‰ C ¡ ª C] ¿ “ ¢” ‟ , Ł C fi S ˛ ¯ Ø ˘ » f ł Œ Ø œ p ˛ p ^ [‟ ƒ Œ ˛ C ł ƒ · › œ p ˛ p ^ [‹ • Ø – ˘ ‟ ¯ « Ø ˘ l ƒ Œ Ø D Œ ß C £ U ^ ⁄ N \ » ˛ – s ` ‰ ° C ' fi o ‰ œ p ˛ p ^ [ ″ › ¨ ' ` ‰ D ¡ ª ″ C £ U ^ ⁄ N \ » ˛ ‰ œ p ˛ p ^ [• K v ‟ Ø ˘ l ƒ Œ Ø D

Q l ¶ £

[1] F , ‰ G , “ ˛ ¶ ' @ B | Œ p Œ « ‹”, æ w _ ¶ , Vol.35, No.11, pp.2283-2290, 1994.

[2] k ″ q, … { T ¡, “ ˛ R [p X p ‰ ˛ \ » ' fi o”, æ w _ ¶ , Vol.38, No.4, pp727-736, 1997.

[3] r · , @, ″ ‟ i, “ K ″ R [p X ' A ‰ ^ ¤ £ U ^ ⁄ N \ » ' fi o @”, æ w _ ¶ , Vol.36, No.11, pp.2584-2596, 1995.

[4] g G ‰, “*N*-gram v p ‰ œ { Œ ⁄ N \ » « ‹”, „ w H w m \ æ H w ¨ † ˘ _ ¶, 1999.

[5] „ c S, ″ ª m Œ, ¿ v ° l, r · , “ œ p ˛ p ^ [' fi o fl ~”, æ w ⁄ æ , 2003-NL-153, pp.113-118, 2003.

[6] ″ ª m Œ, “ p œ ˛ f [^ x [X » ”, u Œ, F fl, \ » v 7 æ N ⁄ v O , 2002.

[7] ¡ Y, “ K ″ R [p X ' d ¶ ¡ ¶ Œ \ ¢ ' fi o”, „ w H w m \ æ H w ¨ † ˘ _ ¶, 2000.