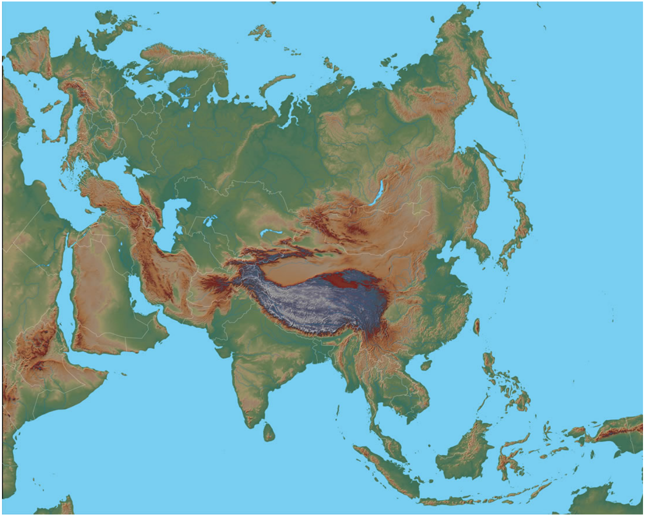
Analogy for Natural Language Processing and Machine Translation

Yves LEPAGE

Waseda University

Contents

France – Japan



Waseda University

- Known for its faculty of literature, its faculty of law, has no faculty of medicine
- Alumni: famous Japanese writers, actors, politicians, CEOs of world-wide companies
- Open to foreign students from its creation
- Ranked n^{o} 1 in Japan for employability

Waseda University

- Known for its faculty of literature, its faculty of law, has no faculty of medicine
- Alumni: famous Japanese writers, actors, politicians, CEOs of world-wide companies
- Open to foreign students from its creation
- Ranked n^{o} 1 in Japan for employability
- Excellent rugby team, 1st or 2nd in Japan university championship

Waseda University

学問の独立 = Independence of scientific and technical knowledge

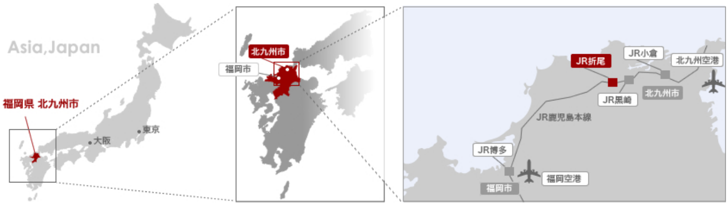
ÔKUMA Sigenobu, founder of Waseda university in 1882, twice prime minister of Japan. Activisit in favour of scientific progress and the enlightenment of people. Eager to welcome foreign students from abroad.

Waseda University

理想の光 = Light of reason

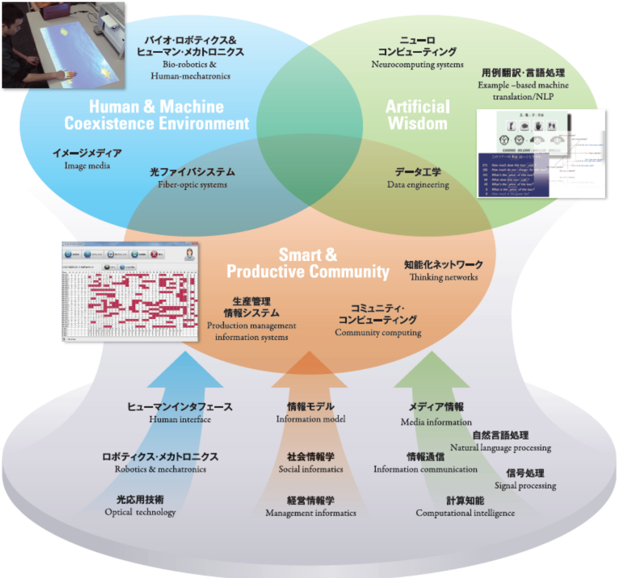
SUGIHARA Tiune, graduated from Waseda university. Saved thousands of Lithuanian and Polish Jews by delivering visas to Japan in violation of the orders of his hierarchy. “Righteous among the nations” after his death in 1985.

Graduate School of Information, Production and Systems (IPS)



Graduate School of Information, Production and Systems (IPS)

Field of Information Architecture



EB(MT/NLP) laboratory

- Machine translation
- Natural language processing
- using example-based methods

- Resources for Chinese–Japanese machine translation (word segmentation, terminology, corpora)
- Alignment methods for machine translation (word-to-word, sub-sentential)
- Proportional analogy on language data (kanji, words, chunks, sentences)

1. Waseda ⊃ IPS ⊃ Information Architecture ⊃ EB(MT/NLP) lab

2. Proportional analogy
   - Examples, definition and usefulness
   - Analogy in morphology
   - Solving analogical equations
   - Machine translation by analogy

Distributional semantics

Word vector representations allow to represent words as vectors [Pantel and Turney, 2010].

What is to a woman in the same way as a king is to a man?

1. *empress*
2. *queen*
3. *princess*

Distributional semantics

Word vector representations allow to represent words as vectors [Pantel and Turney, 2010].

What is to a woman in the same way as a king is to a man?

1.
2. *queen*
3.

Distributional semantics

Word vector representations allow to represent words as vectors [Pantel and Turney, 2010].

What is to a woman in the same way as a king is to a man?

1.
2. *queen*
3.

Word vector representations (Word2Vec by [Mikolov, 2013] or GloVe [Pennington et al., 2014])

$$man : king :: woman : x \quad \Rightarrow \quad x =$$

Distributional semantics

Word vector representations allow to represent words as vectors [Pantel and Turney, 2010].

What is to a woman in the same way as a king is to a man?

1.
2. *queen*
3.

Word vector representations (Word2Vec by [Mikolov, 2013] or GloVe [Pennington et al., 2014])

$$man : king :: woman : x \quad \Rightarrow \quad x = queen$$

Distributional semantics

Word vector representations allow to represent words as vectors [Pantel and Turney, 2010].

What is to a woman in the same way as a king is to a man?

1.
2. *queen*
3.

Word vector representations (Word2Vec by [Mikolov, 2013] or GloVe [Pennington et al., 2014])

$$man : king :: woman : x \quad \Rightarrow \quad x = queen$$

$$\overrightarrow{\text{queen}}$$

Distributional semantics

Word vector representations allow to represent words as vectors [Pantel and Turney, 2010].

What is to a woman in the same way as a king is to a man?

1.
2. *queen*
3.

Word vector representations (Word2Vec by [Mikolov, 2013] or GloVe [Pennington et al., 2014])

$$man : king :: woman : x \quad \Rightarrow \quad x = queen$$

$$\overrightarrow{\text{queen}} \approx \overrightarrow{king} - \overrightarrow{man} + \overrightarrow{woman}$$

Analogy

foot : *shoe* :: *hand* :

Analogy

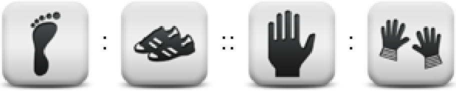
foot : *shoe* :: *hand* : *glove*

Analogy

foot : *shoe* :: *hand* : *glove*

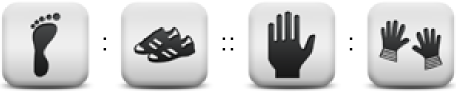
Analogy

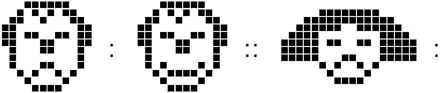
foot : *shoe* :: *hand* : *glove*



Analogy

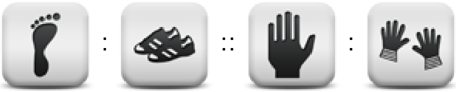
foot : *shoe* :: *hand* : *glove*

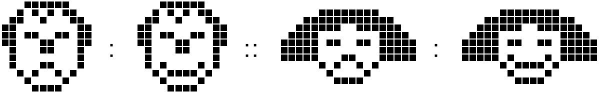




Analogy

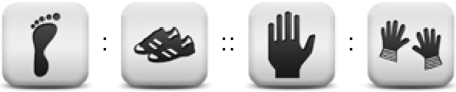
foot : *shoe* :: *hand* : *glove*

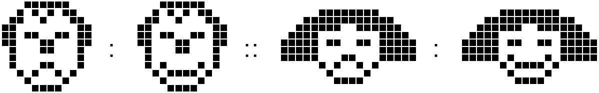


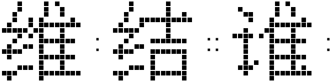


Analogy

foot : *shoe* :: *hand* : *glove*

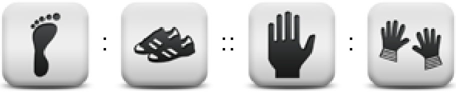


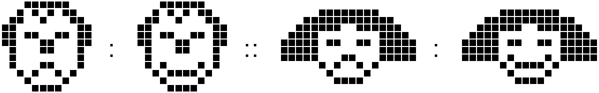


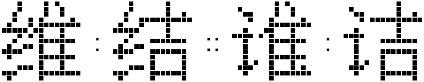


Analogy

foot : *shoe* :: *hand* : *glove*

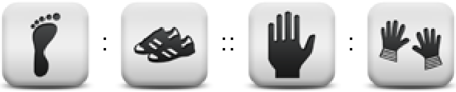


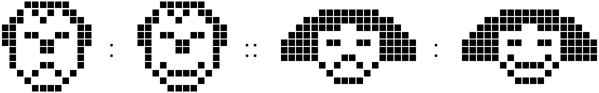


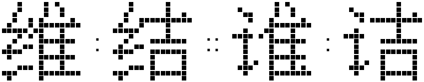


Analogy

foot : *shoe* :: *hand* : *glove*



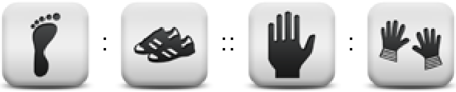


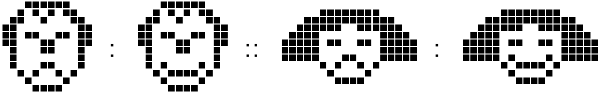


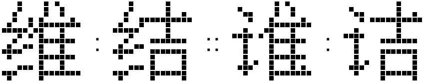
01000000 : 00110000 :: 00011100 :

Analogy

foot : *shoe* :: *hand* : *glove*



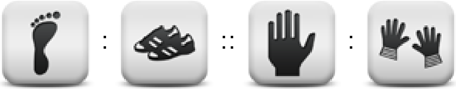


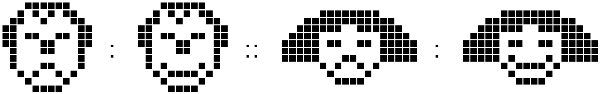


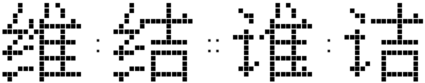
01000000 : 00110000 :: 00011100 : 00001111

Analogy

foot : *shoe* :: *hand* : *glove*





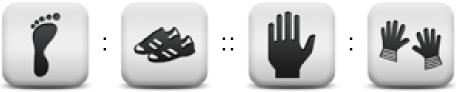


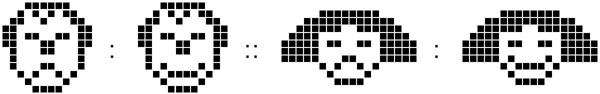
01000000 : 00110000 :: 00011100 : 00001111

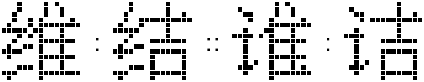
take : *taking* :: *elaborate* :

Analogy

foot : *shoe* :: *hand* : *glove*



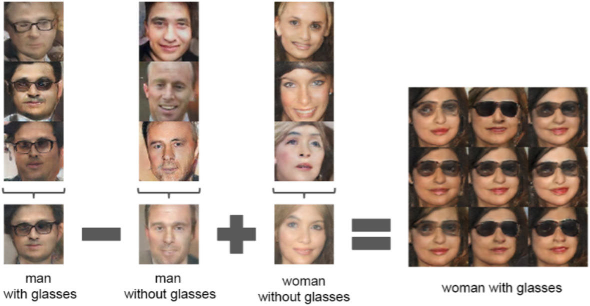




01000000 : 00110000 :: 00011100 : 00001111

take : *taking* :: *elaborate* : *elaborating*

Analogy



Picture from Radford et al., Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, ArXiv, 2015.

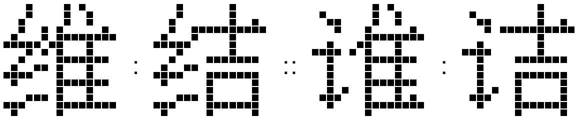
Usefulness in learning [Matsushita, 2013]

Hypothesis: correspondences on the three levels of form, sound and meaning should help in remembering.

洛	:	珞	::	渝	:	瑜
river	:	*accessory*	::	*river*	:	*accessory*
/lùo/	:	*/lùo/*	::	*/yú/*	:	*/yú/*
Shaanxi	:	*necklace*	::	*Chongqing*	:	*gem, jewel*

Matsushita (2013) showed that remembering groups of 4 hanzi is eased when graphical analogies can be found between them.

Chinese characters = pixel images



Production of analogical clusters (1/4) [Lepage, 2014]

$$\begin{pmatrix}0\\0\\0\\0\\0\end{pmatrix} \quad \begin{pmatrix}1\\0\\0\\1\\0\end{pmatrix} \quad \begin{pmatrix}1\\1\\0\\2\\0\end{pmatrix} \quad \begin{pmatrix}2\\2\\2\\0\\2\end{pmatrix} \quad \begin{pmatrix}3\\2\\2\\1\\2\end{pmatrix}$$

$$A \quad B \quad C \quad D \quad E$$

Production of analogical clusters (2/4)

[*A*–*E*]:none

level 1:	[*A*–*A*]:0	[*B*–*C*]:1		[*D*–*D*]:2	[*E*–*E*]:3
level 2:	[*A*–*A*]:0	[*B*–*B*]:0	[*C*–*C*]:1	[*D*–*D*]:2	[*E*–*E*]:2
level 3:	[*A*–*A*]:0	[*B*–*B*]:0	[*C*–*C*]:0	[*D*–*D*]:2	[*E*–*E*]:2
level 4:	[*A*–*A*]:0	[*B*–*B*]:1	[*C*–*C*]:2	[*D*–*D*]:0	[*E*–*E*]:1
level 5:	[*A*–*A*]:0	[*B*–*B*]:0	[*C*–*C*]:0	[*D*–*D*]:2	[*E*–*E*]:2

Production of analogical clusters (3/4)

Same ratio = same difference on each level
Level 1:

	[A–A]:0	[B–C]:1	[D–D]:2	[E–E]:3
[A–A]:0		1	2	3
[B–C]:1		0	1	2
[D–D]:2				1
[E–E]:3				

Production of analogical clusters (3/4)

Same ratio = same difference on each level
Level 2:

	$[A–A]$:0	$[B–B]$:0	$[C–C]$:1	$[D–D]$:2	$[E–E]$:2
$[A–A]$:0		1,0	1,1	2,2	
$[B–B]$:0			0,1	1,2	2,2
$[C–C]$:1				1,1	2,1
$[D–D]$:2					1,0
$[E–E]$:2					

Production of analogical clusters (3/4)

Same ratio = same difference on each level
Level 3:

	$[A–A]$:0	$[B–B]$:0	$[C–C]$:0	$[D–D]$:2	$[E–E]$:2
$[A–A]$:0		1,0,0	1,1,0	2,2,2	
$[B–B]$:0					2,2,2
$[C–C]$:0				1,1,2	
$[D–D]$:2					1,0,0
$[E–E]$:2					

Production of analogical clusters (4/4)

Result:

The method reduces a problem with a complexity of $O(n^4)$ to an exploration in $O(n^2)$.

It enumerates all series of pairs of objects (represented by their vector) with the same ratio, i.e., analogical clusters.

$$\begin{array}{c} W_1^1 : W_1^2 \\ W_2^1 : W_2^2 \\ \vdots \quad \vdots \\ W_n^1 : W_n^2 \end{array} \quad \stackrel{\Delta}{\Longleftrightarrow} \quad \forall (i,j) \in \{1, \ldots, n\}^2, \qquad W_i^1 : W_i^2 :: W_j^1 : W_j^2 \tag{1}$$

Definition of an analogical cluster.

Analogical clusters of Chinese char. [Lepage, 2014]

倔:掘
恨:抿
怕:拍
惜:措
快:抉
怜:拎
惦:掂
俸:捧

诘:结
调:绸
谝:编
谁:维

捂:梧
抗:杭
拮:桔

偏:惆
谝:调
编:绸

诅:祖
诈:祚

铂:珀
锂:理

冂:冋
口:回
匚:叵

另:另
佘:余

1. Waseda $\supset$ IPS $\supset$ Information Architecture $\supset$ EB(MT/NLP) lab

2. Proportional analogy
   - Examples, definition and usefulness
   - Analogy in morphology
   - Solving analogical equations
   - Machine translation by analogy

Paradigm tables

inf.	:	preterit	:	pres. part.	:	past part.
to be	:	*was*	:	*being*	:	*been*
to take	:	*took*	:	*taking*	:	*taken*
to look	:	*looked*	:	*looking*	:	*looked*
to walk	:	*walked*	:	*walking*	:	*walked*

Figure: A paradigm table with exponents (header) and lemmas (first column).

Anto memakan nasi dan meminum air. Nasi itu dibeli di pasar. Di pasar, Anto melihat mainan. Anto senang main bola. Setelah main, Anto suka minum es dan makan cilok. Makanan dan minuman itu juga dia beli di pasar. Es dan cilok memang enak dimakan dan diminum selesai olahraga.

Figure: A text in Indonesian.

*air anto **beli** bola cilok dan di dia **dibeli** **dimakan** **diminum** enak es itu juga **main** **mainan** **makan** **makanan** melihat **memakan** memang **meminum** **minum** **minuman** nasi olahraga pasar selesai senang setelah suka*

Figure: List of words extracted from the previous text.

Analogical grids [Fam and Lepage, 2016a]

makan	:	*dimakan*	:	*memakan*	:	*makanan*
minum	:	*diminum*	:	*meminum*	:	*minuman*
main	:		:		:	*mainan*
beli	:	*dibeli*	:		:	

Figure: An analogical grid obtained from the set of previous words.

Analogical grids [Fam and Lepage, 2016a]

makan	:	*dimakan*	:	*memakan*	:	*makanan*
minum	:	*diminum*	:	*meminum*	:	*minuman*
main	:		:		:	*mainan*
beli	:	*dibeli*	:		:	

Figure: An analogical grid obtained from the set of previous words.

Analogical grids [Fam and Lepage, 2016a]

makan	:	*dimakan*	:	*memakan*	:	*makanan*
minum	:	*diminum*	:	*meminum*	:	*minuman*
main	:		:		:	*mainan*
beli	:	*dibeli*	:		:	

Figure: An analogical grid obtained from the set of previous words.

Analogical grids [Fam and Lepage, 2016a]

makan	:	*dimakan*	:	*memakan*	:	*makanan*
minum	:	*diminum*	:	*meminum*	:	*minuman*
main	:		:		:	*mainan*
beli	:	*dibeli*	:		:	

Figure: An analogical grid obtained from the set of previous words.

$$\begin{array}{l} W_1^1 : W_1^2 : \cdots : W_1^m \\ W_2^1 : W_2^2 : \cdots : W_2^m \\ \vdots \quad \vdots \qquad\quad \vdots \\ W_n^1 : W_n^2 : \cdots : W_n^m \end{array} \quad \overset{\Delta}{\Longleftrightarrow} \quad \begin{array}{l} \forall (i,k) \in \{1,\ldots,n\}^2, \\ \forall (j,l) \in \{1,\ldots,m\}^2, \end{array} \quad W_i^j : W_i^l :: W_k^j : W_k^l \tag{2}$$

Definition of an analogical grid. Caution: there may be empty cells.

$$A : B :: C : D \quad \stackrel{\Delta}{\Longleftrightarrow} \quad \begin{cases} A : B = C : D \\ A : C = B : D \end{cases} \tag{3}$$

Definition of an analogy.

$$A : B \quad \triangleq \quad \begin{pmatrix} |A|_a - |B|_a \\ \vdots \\ |A|_z - |B|_z \\ \mathrm{d}(A, B) \end{pmatrix} \tag{4}$$

Definition of a ratio.

$$A : B \quad \triangleq \quad \begin{pmatrix} |A|_a - |B|_a \\ \vdots \\ |A|_z - |B|_z \\ \mathrm{d}(A, B) \end{pmatrix} \tag{4}$$

Definition of a ratio.

$$\textit{makan} : \textit{makanan} \quad =$$

$$A : B \quad \triangleq \quad \begin{pmatrix} |A|_a - |B|_a \\ \vdots \\ |A|_z - |B|_z \\ \mathrm{d}(A, B) \end{pmatrix} \tag{4}$$

Definition of a ratio.

$$makan : makanan \quad = \quad \left(\begin{matrix} \text{(a) } 2 - 3 \\ \\ \\ \\ \\ \end{matrix} \right.$$

$$A : B \quad \triangleq \quad \begin{pmatrix} |A|_a - |B|_a \\ \vdots \\ |A|_z - |B|_z \\ \mathrm{d}(A, B) \end{pmatrix} \tag{4}$$

Definition of a ratio.

$$\textit{makan} : \textit{makanan} \quad = \quad \left(\begin{array}{c} \text{(a) } 2 - 3 \\ \vdots \\ \text{(n) } 1 - 2 \\ \\ \\ \\ \end{array} \right.$$

$$A : B \quad \triangleq \quad \begin{pmatrix} |A|_a - |B|_a \\ \vdots \\ |A|_z - |B|_z \\ \mathrm{d}(A, B) \end{pmatrix} \tag{4}$$

Definition of a ratio.

$$makan : makanan \quad = \quad \left(\begin{array}{c} \text{(a) } 2 - 3 \\ \vdots \\ \text{(n) } 1 - 2 \\ \vdots \\ \text{(z) } 0 - 0 \end{array} \right.$$

$$A : B \quad \triangleq \quad \begin{pmatrix} |A|_a - |B|_a \\ \vdots \\ |A|_z - |B|_z \\ \mathrm{d}(A, B) \end{pmatrix} \tag{4}$$

Definition of a ratio.

$$\mathit{makan} : \mathit{makanan} \quad = \quad \begin{pmatrix} \text{(a) } 2 - 3 \\ \vdots \\ \text{(n) } 1 - 2 \\ \vdots \\ \text{(z) } 0 - 0 \\ \mathrm{d}(\mathit{makan}, \mathit{makan}\textcolor{red}{\mathit{an}}) \end{pmatrix}$$

$$A : B \quad \triangleq \quad \begin{pmatrix} |A|_a - |B|_a \\ \vdots \\ |A|_z - |B|_z \\ \mathrm{d}(A, B) \end{pmatrix} \tag{4}$$

Definition of a ratio.

$$\textit{makan} : \textit{makanan} \quad = \quad \begin{pmatrix} \text{(a) } 2 - 3 \\ \vdots \\ \text{(n) } 1 - 2 \\ \vdots \\ \text{(z) } 0 - 0 \\ \mathrm{d}(\textit{makan}, \textit{makanan}) \end{pmatrix} \quad = \quad \begin{pmatrix} -1 \\ \vdots \\ -1 \\ \vdots \\ 0 \\ 2 \end{pmatrix}$$

Types of analogies not covered by this definition:

$$abc : def :: ghi : jkl$$

Types of analogies **not** covered by this definition:

$$abc : def :: ghi : jkl$$

あいう : えおか :: きくけ :

Types of analogies not covered by this definition:

$$abc : def :: ghi : jkl$$

あいう : えおか :: きくけ : こさし

Types of analogies not covered by this definition:

abc : *def* :: *ghi* : *jkl*
あいう : えおか :: きくけ : こさし
man : *woman* :: *king* : *queen*
fish : *fins* :: *bird* : *wings*

Types of analogies not covered by this definition:

abc : *def* :: *ghi* : *jkl*
あいう : えおか :: きくけ : こさし
man : *woman* :: *king* : *queen*
fish : *fins* :: *bird* : *wings*

burung : *burung-burung* :: *kucing* : *kucing-kucing*
'a bird' : *'birds'* :: *'a cat'* : *'cats'*
λύω : λέλυκα :: δακρύω : δεδάκρυκα
'I unlash' : *'I unlashed'* :: *'I cry'* : *'I cried'*
高い : 高ければ高いほど. . . :: よい : よければよいほど. . .
'high' : *'the higher, the. . . '* :: *'good'* : *'the better, the. . . '*

Producing analogical grids (1/2)

- Produce analogical clusters with words represented as vectors
  (number of occurrences of characters)
- Check for equality of edit distances
- Produce analogical grids by combining compatible clusters either vertically or horizontally
  (next slide)

Producing analogical grids (2/2)

- Sort list of analogical clusters by length.
- Take longest analogical cluster as a new analogical grid.
- For each analogical cluster of the list:
  - If it can be added (vertically or horizontally) to the analogical grid [check density threshold too] then
    - Add it to the table.
    - Remove it from the list of analogical clusters.
- Scan the list several times until no more cluster can be added into the table.
- Redo from second step until all analogical clusters are used up.

Tools downloadable at:

`http://lepage-lab.ips.waseda.ac.jp/`
`kakenhi-2-tools-released`

Programs:

- `solveanalogy`, `verifanalogy`,
- `Words2Clusters`, `Words2Grids`,
- etc.

Language	# tokens (N)	# types (V)	Length of types avg$\pm$std. dev.	# grids	Time (h:min)
English	792,074	12,498	7.03 ± 2.18	12,855	45
Indonesian	648,606	15,641	7.84 ± 2.63	25,752	2:04
Modern Greek	706,771	36,786	8.49 ± 2.49	69,173	11:03
Russian	560,524	47,226	8.26 ± 2.73	60,035	10:34

Table: Number of analogical grids produced with the time needed to produce them on the New Testament, corpus[2] collected by Christodoulopoulos as a continuation of work by Resnik (1999).

[1] http://homepages.inf.ed.ac.uk/s0787820/bible/

[2] http://homepages.inf.ed.ac.uk/s0787820/bible/

Saturation of analogical grids

$$\text{Saturation} = \frac{\text{Number of non-empty cells}}{\text{Total number of cells = Size}} \times 100\% \qquad (5)$$

Saturation of analogical grids

$$\text{Saturation} = \frac{\text{Number of non-empty cells}}{\text{Total number of cells = Size}} \times 100\% \tag{5}$$

makan	:	*dimakan*	:	*memakan*	:	*makanan*
minum	:	*diminum*	:	*meminum*	:	*minuman*
main	:		:		:	*mainan*
beli	:	*dibeli*	:		:	

Figure: An analogical grid with a saturation of $\frac{12}{16} \times 100\% = \frac{3}{4} \times 100\% = 75\%$

Size, saturation, and number of analogical grids

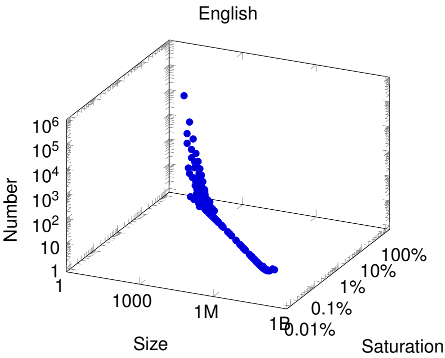


Figure: Number of analogical grids obtained against their size and saturation in English. Algorithmic scale on the three axes.

Size vs. saturation of analogical grids

$$\log(\text{saturation}) = a \times \log(\text{size}) + b$$

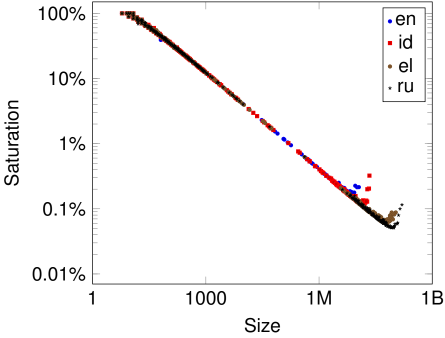


Figure: Saturation of paradigm tables against size in each language

$\log(\text{saturation}) = a \times \log(\text{size}) + b$ independent of language, size and genre?

Table: Linear coefficients for each language; and for different sizes and different genres in English.

Language	Data and size	Range for saturation [0%,100 %] *a*	[0%,100 %] *b*	[50%,100 %] *a*	[50%,100 %] *b*
English	Bible 100.0 %	-0.480	0.510	-0.366	0.332
	50.0 %	-0.479	0.507	-0.372	0.343
	25.0 %	-0.476	0.499	-0.368	0.336
	12.5 %	-0.474	0.491	-0.361	0.323
	Europarl (same size as Bible)	-0.481	0.516	-0.365	0.333
Indonesian	Bible 100.0 %	-0.481	0.518	-0.371	0.343
Modern Greek	"	-0.479	0.514	-0.369	0.342
Russian	"	-0.482	0.520	-0.370	0.342

Use of analogical grids

show	:	shows	:	showing	:	showed
walk	:	walks	:	walking	:	walked
open	:	opens	:	opening	:	
study	:		:	studying	:	

Figure: An analogical grid in English

Generating new words

show	:	shows	:	showing	:	showed
walk	:	walks	:	walking	:	walked
open	:	opens	:	opening	:	
study	:	studys	:	studying	:	studyed

Figure: An analogical grid in English

Explaining unseen words

show	:	shows	:	showing	:	showed
walk	:	walks	:	walking	:	walked
open	:	opens	:	opening	:	opened
study	:		:	studying	:	

Figure: An analogical grid in English

Examples of words explained by analogical grids

Reasoning and *glorifying* are two words which appear in Luke but not in Matthew. They are explained by the analogical grids built from the words in Matthew.

Examples of words explained by analogical grids

Reasoning and *glorifying* are two words which appear in Luke but not in Matthew. They are explained by the analogical grids built from the words in Matthew.

enter	:	entering	:	entered
reason	:	reasoning	:	reasoned
mourn	:	mourning	:	mourned
open	:		:	opened

Examples of words explained by analogical grids

Reasoning and *glorifying* are two words which appear in Luke but not in Matthew. They are explained by the analogical grids built from the words in Matthew.

enter :	entering :	entered	marry :	marrying :	married
reason :	reasoning :	reasoned	deny :	:	denied
mourn :	mourning :	mourned	cry :	crying :	cried
open :	:	opened	glorify :	glorifying :	glorified

Languages and texts used

Language	Training set (Matthew)			Test set (Luke)		
	Number of tokens	Number of types	Type-token ratio (%)	Number of tokens	Number of types	Type-token ratio (%)
English	23,726	2,098	8.8	25,987	2,370	9.1
Indonesian	22,375	2,450	10.9	23,623	2,650	11.2
Nahuatl	23,222	3,833	16.5	24,060	4,096	17.0
Greek	20,438	3,819	18.7	21,856	4,367	20.0
Somali	20,375	3,967	19.5	21,535	4,244	19.7
Chinese	18,350	4,030	22.0	19,956	4,488	22.5
Achuar	22,470	5,349	23.8	23,177	5,609	24.2
Finnish	17,331	4,467	25.8	18,804	5,003	26.6
Quichua	15,038	4,066	27.0	16.332	4,249	26.0
Swahili	16,851	3,926	23.3	18,467	4,411	23.9
Xhosa	14,505	5,580	38.5	15,537	6,265	40.3
Telugu	13,083	6,066	46.4	14,404	6,747	46.8

Explaining unseen words using analogical grids

	Productive analogical grids (Matthew)			Unseen words (Luke)		
Language	Total number	Avg size	Avg density (%)	Total number	Explained	Ratio (%)
English	587	49.5	58.3	858	75	8.7
Indonesian	790	48.3	57.8	940	126	13.4
Chinese	220	88.4	55.9	2,497	193	7.7
Finnish	2,147	49.7	57.6	2,597	331	12.7
Nahuatl	512	67.9	57.1	2,143	296	13.8
Greek	793	64.1	57.7	2,238	352	15.7
Somali	2,078	61.8	55.1	1,929	392	20.3
Swahili	2,067	53.6	56.5	2,381	430	18.1
Telugu	557	74.4	56.0	4,485	459	10.2
Xhosa	3,501	60.2	55.2	3,807	734	19.3
Achuar	11,349	49.1	53.4	2,801	748	26.7
Quichua	4,478	59.3	55.0	2,170	900	41.5

Explaining unseen words on the levels of form, morphology and distributional semantics

Form:	*makan* : *makanan* :: *minum* : *minuman*
Morphological representation:	*makan_VB* : *makan+an_NN* :: *minum_VB* : *minum+an_NN*
Semantic representation:	$\overrightarrow{makanan} - \overrightarrow{makan} + \overrightarrow{minum} \approx \overrightarrow{minuman}$

Figure: Confirming an analogy on different levels of representation for the word *minuman*.

Number of unseen words explained

Surface form	Morphology	Distributional semantics	Total	
✓			1,249	98 %
✓	✓		1,010	79 %
✓		✓	791	62 %
✓	✓	✓	724	57 %

Table: Number of unseen words that can be explained on three different levels. Ten-fold cross-validation on the BPPT (nearly half a million tokens, 27,000 types)

Examples of unseen words explained

F	M	S	Number	Examples	English translation
✓	×	×	172	*ilustrasi* *terenggut* *Montolivo*	‘illustration‘ ‘wrenched‘ person’s name
✓	✓	×	286	*disewakan* *bercampur* *menyepakatinya*	‘for rent‘ ‘mixed‘ ‘to agree‘
✓	×	✓	67	*endoplasma* *perfeksionis* *radjawali*	‘endoplasm‘ ‘perfectionist‘ name of a kind of bird
✓	✓	✓	724	*persilangan* *terkoordinasi* *pembelajaran*	‘crossing‘ ‘coordinated‘ ‘learning‘

Table: Examples of unseen words explained or not on each level of representation: surface form (F), morphological representation (M), and distributional semantic representation (S).

Standard techniques [Lepage, 1998]

First algorithm based on the use of edit distance traces. Very fast.

(Chinese)	科学 : 科学家 :: 政治 : 政治家
(French)	*recevoir* : *j'ai reçu* :: *percevoir* : *j'ai perçu*
(German)	*sprechen* : *ihr sprächet* :: *nehmen* : *ihr nähmet*
(Hebrew)	*mélex* : *mlaxím* :: *rések* : *rsakím*
(Malay)	*kawan* : *mengawani* :: *keliling* : *mengelilingi*
(Polish)	*stworzyć* : *stwarzać* :: *rozmnożyć się* : *rozmnażać się*
(formal)	*abc* : *abcabc* :: *abcabcabc* : *abcabcabcabc*
(formal)	*ab* : *aabb* :: *aaaaaabbbbbb* : *aaaaaaabbbbbbb*

Standard techniques [Lepage, 2017]

Recent algorithm based on the use of edit distance traces.

- Data from Task 1 of Track 1 of SIGMORPHON 2016 Shared Task: Morphological Reinflection in 10 different languages.[3]
- Analogy questions built by extracting all analogies of form, and filtering by morphological features. Each analogy yields four different analogy questions.[4]

$$
\begin{array}{rcl}
\textit{alterado} : \textit{alterada} :: \textit{adeudados} : x & \Rightarrow & x = \textit{adeudadas} \\
\textit{alterada} : \textit{alterado} :: \textit{adeudadas} : x & \Rightarrow & x = \textit{adeudados} \\
\textit{adeudadas} : \textit{adeudados} :: \textit{alterada} : x & \Rightarrow & x = \textit{alterado} \\
\textit{adeudados} : \textit{adeudadas} :: \textit{alterado} : x & \Rightarrow & x = \textit{alterada}
\end{array}
$$

[3] All files `<language>-task1-train` from https://github.com/ryancotterell/sigmorphon2016/tree/master/data/

[4] Different from the task proposed in SIGMORPHON Shared Task.

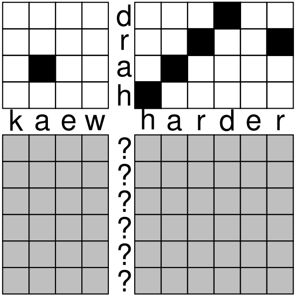
Standard techniques

Language	Number of analogical equations	% of correct answers
Arabic	381,132	94
Finnish	3,076	95
Georgian	7,256,156	87
German	349,796	91
Hungarian	15,157,368	94
Maltese	10,000	97
Navajo	18,588,020	97
Russian	66,672	99
Spanish	95,564	95
Turkish	729,092	86
Total	42,636,876	94

Table: Solving analogical equations extracted from all training data of Task 1 of Track 1 from SIGMORPHON 2016 Shared Task.

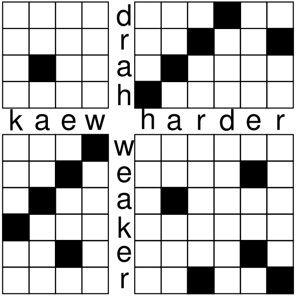
Neural networks for analogy

$$hard : harder :: weak : x \quad \Rightarrow \quad x = \quad ?$$

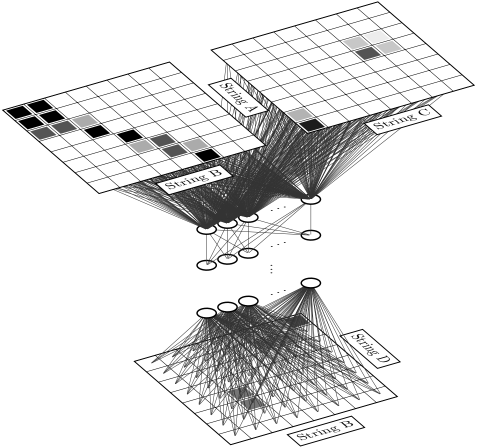


Neural networks for analogy

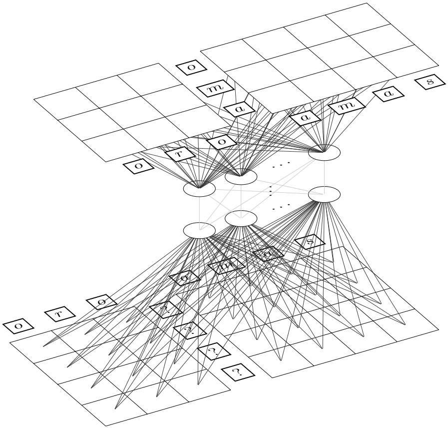
$$\textit{hard} : \textit{harder} :: \textit{weak} : x \quad \Rightarrow \quad x = \textit{weaker}$$



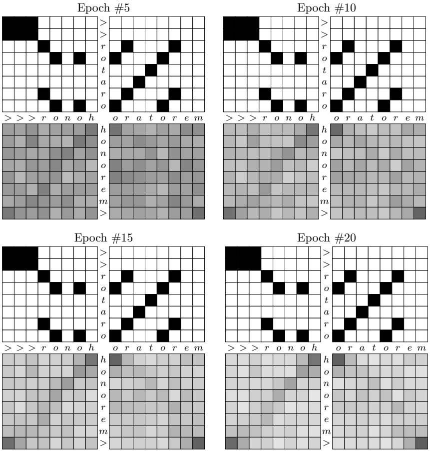
First model [Kaveeta and Lepage, 2016]



Second model [Kaveeta, 2017]



$$amo : oro :: amas : x \quad \Rightarrow \quad x = oras$$



$$\textit{orator} : \textit{honor} :: \textit{oratorem} : x \quad \Rightarrow \quad x = \textit{honorem}$$

Data used

oratorem : *orator* :: *honorem* : *honor*
huzila : *huzāl* :: *ṣudi'a* : *ṣudā'*
setzen : *setzte* :: *lachen* : *lachte*
inné : *nées* :: *indu* : *dues*
biorąc : *bierzesz* :: *piorąc* : *pierzesz*
tinggal : *ketinggalan* :: *duduk* : *kedudukan*
aa : *ab* :: *ba* : *bb*
ab : *aabb* :: *aaaaaaabbbbbbb* : *aaaaaaaabbbbbbbb*
abc : *aabbcc* :: *aaabbbccc* : *aaaabbbbcccc*
aab : *aaaabb* :: *aaaaaabbb* : *aaaaaaaabbbb*
aaaabbbbcccc : *aaabbbccc* :: *aabbcc* : *abc*

Results for the first model (90 % train, 10 % test)

		# of hyper parameters	Train time (m:s)	Train loss (MSE)	Test loss (MSE)	Accuracy (%)
Alignment matrices size	2 × 2	1,668	4.07	0.009	0.005	1.73
	4 × 4	6,288	4.53	0.013	0.008	16.75
	8 × 8	24,768	5.34	0.017	0.010	67.18
	16 × 16	98,688	7.12	0.024	0.017	79.10
	32 × 32	394,368	14.03	0.035	0.026	**84.11**
Re-sampling methods	NN	98,688	6.56	0.039	0.031	67.88
	Bilinear	98,688	7.10	0.015	0.010	72.71
	Bicubic	98,688	6.40	0.019	0.012	78.24
	Proposed	98,688	7.12	0.024	0.017	**79.10**
Filtering methods	None	98,688	6.28	0.056	0.044	77.72
	Morph	98,688	6.39	0.040	0.031	76.68
	Weight	98,688	7.32	0.034	0.025	79.45
	Both	98,688	7.12	0.024	0.017	**80.48**
Number of hidden nodes	128	98,688	7.12	0.024	0.017	80.83
	256	197,120	8.21	0.022	0.015	82.38
	512	393,984	9.51	0.021	0.017	83.07
	1024	787,712	13.28	0.019	0.012	**85.84**
Number of hidden layers	1	98,688	7.12	0.024	0.017	80.66
	2	115,200	9.55	0.023	0.015	84.44
	3	131,712	11.22	0.023	0.014	86.36
	4	148,224	11.54	0.023	0.014	**87.56**

Results for the second model (5 % train, 95 % test)

Techniques	Model	Mat. dim.	Accuracy (%)
Nearest neighbour	Single	32 x 32	11.68
Linear	Single	32 x 32	14.19
Bilinear	Single	32 x 32	15.82
Bicubic	Single	32 x 32	15.76
Generation	Multiple	Variables	**55.86**
Generation (Small training set)[5]	Multiple	Variables	36.29

[5] 10 analogies

Third model? [Zhao, 2018]?

- Use of generative adversarial networks (GAN)?
- Use of Siamese neural networks?

1. Waseda $\supset$ IPS $\supset$ Information Architecture $\supset$ EB(MT/NLP) lab

2. Proportional analogy
   - Examples, definition and usefulness
   - Analogy in morphology
   - Solving analogical equations
   - Machine translation by analogy

Suppose that the sentence 濃い紅茶が飲みたい。

translates into *I would like a cup of strong tea.*

How to translate

the sentence 濃いコーヒーが飲みたい。

in the same way?

Suppose that the sentence	濃い紅茶が飲みたい。
translates into	*I would like a cup of strong tea.*
In the same way	
the sentence	濃いコーヒーが飲みたい。
translates into	*I would like a cup of strong coffee.*

Suppose that the sentence	濃い紅茶が飲みたい。
translates into	*I would like a cup of strong tea.*
In the same way,	
the sentence	濃いコーヒーが飲みたい。
translates into	*I would like a cup of strong coffee.*
because	
the word	紅茶
is to the word	*tea*
in the same way as	
the word	コーヒー
is to the word	*coffee*

The sentence
is to the sentence
in the same way as
the sentence
is to the sentence
because
the word
is to the word
in the same way as
the word
is to the word

濃い紅茶が飲みたい。
I would like a cup of strong tea.

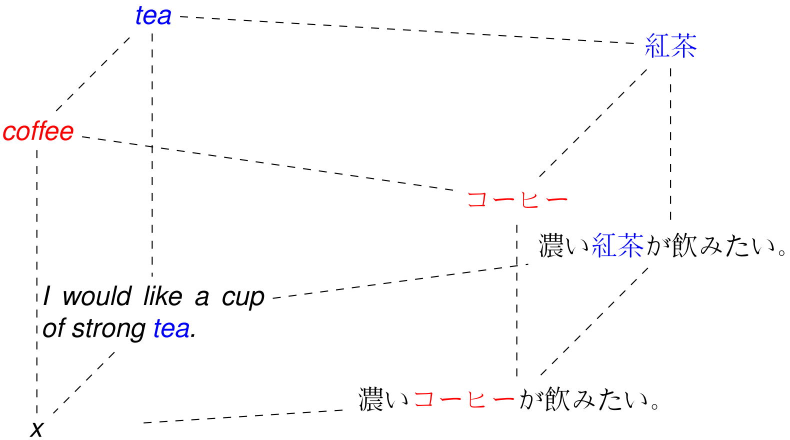
濃いコーヒーが飲みたい。
I would like a cup of strong coffee.

紅茶
tea

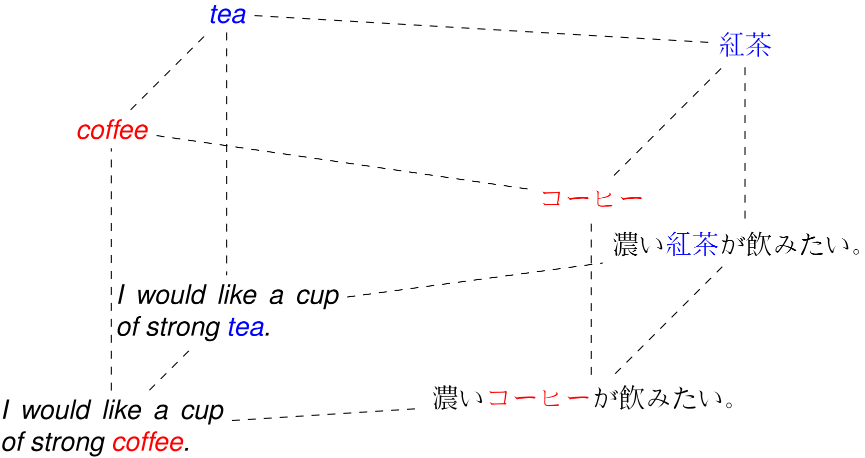
コーヒー
coffee

The sentence	濃い紅茶が飲みたい。
is to the word	紅茶
in the same way as	
the sentence	濃いコーヒーが飲みたい。
is to the word	コーヒー
because	
the sentence	*I would like a cup of strong tea.*
is to the word	*tea*
in the same way as	
the sentence	*I would like a cup of strong coffee.*
is to the word	*coffee*

Principle


tea
紅茶
coffee
コーヒー
濃い紅茶が飲みたい。
I would like a cup of strong tea.
濃いコーヒーが飲みたい。
x

Dictionary / Sentences



Identical contexts

May I have tea, please?

紅茶をください。

May I have coffee, please?

コーヒーをください。

濃い紅茶が飲みたい。

I would like a cup of strong tea.

I would like a cup of strong coffee.

濃いコーヒーが飲みたい。

Different contexts

May I have some tea, please?

紅茶をください。

May I have a cup of coffee?

コーヒーを一杯ください。

濃い紅茶が飲みたい。

I'd like some strong tea, please.

I'd like a cup of strong coffee.

濃いコーヒーが一杯飲みたい。

General case

May I have some tea, please?

紅茶をください。

May I have a cup of coffee?

コーヒーを一杯ください。

濃い紅茶が飲みたい。

I'd like some strong tea, please.

濃いコーヒーが一杯飲みたい。

I'd like a cup of strong coffee.

In Prolog:

% database of facts (bicorpus)

$\texttt{translation}(s_1,\widehat{s_1})\texttt{.}$
$\texttt{translation}(s_2,\widehat{s_2})\texttt{.}$
$\vdots$
$\texttt{translation}(s_n,\widehat{s_n})\texttt{.}$

% translation program proper

$\texttt{translation}(D,\widehat{D})$:–
    $\texttt{translation}(A,\widehat{A})$,
    $\texttt{translation}(B,\widehat{B})$,
    $\texttt{analogy}(A,B,C,D)$,
    $\texttt{translation}(C,\widehat{C})$,
    $\texttt{analogy}(\widehat{A},\widehat{B},\widehat{C},\widehat{D})$,
    $\texttt{assert}(\texttt{translation}(D,\widehat{D}))$.

% translation program proper

translation(d,$\widehat{D}$) :– $\leftarrow$

    translation(A,$\widehat{A}$),
    translation(B,$\widehat{B}$),
    analogy(A,B,C,d),
    translation(C,$\widehat{C}$),
    analogy($\widehat{A}$,$\widehat{B}$,$\widehat{C}$,$\widehat{D}$),
    assert(translation(d,$\widehat{D}$)).

% translation program proper

```
translation(d,D̂) :–
    translation(a,â),
    translation(B,B̂),
    analogy(a,B,C,d),
    translation(C,Ĉ),
    analogy(â,B̂,Ĉ,D̂),
    assert(translation(d,D̂)).
```

$\leftarrow$

% translation program proper

$\mathtt{translation}(d, \widehat{D})$:–
$\quad \mathtt{translation}(a, \widehat{a}),$
$\quad \mathtt{translation}(b, \widehat{b}),$ $\leftarrow$
$\quad \mathtt{analogy}(a, b, C, d),$
$\quad \mathtt{translation}(C, \widehat{C}),$
$\quad \mathtt{analogy}(\widehat{a}, \widehat{b}, \widehat{C}, \widehat{D}),$
$\quad \mathtt{assert}(\mathtt{translation}(d, \widehat{D})).$

% translation program proper

$\texttt{translation}(d,\widehat{D})$:–
    $\texttt{translation}(a,\widehat{a})$,
    $\texttt{translation}(b,\widehat{b})$,
    $\texttt{analogy}(a,b,C,d)$, $\leftarrow$
    $\texttt{translation}(C,\widehat{C})$,
    $\texttt{analogy}(\widehat{a},\widehat{b},\widehat{C},\widehat{D})$,
    $\texttt{assert}(\texttt{translation}(d,\widehat{D}))$.

% translation program proper

```
translation(d,D̂) :-
    translation(a,â),
    translation(b,b̂),                    ←
    analogy(a,b,C,d),
    translation(C,Ĉ),
    analogy(â,b̂,Ĉ,D̂),
    assert(translation(d,D̂)).
```

% translation program proper

```
translation(d,D̂) :–
    translation(a,â),
    translation(b,b̂),
    analogy(a,b,C,d),            ←
    translation(C,Ĉ),
    analogy(â,b̂,Ĉ,D̂),
    assert(translation(d,D̂)).
```

% translation program proper

```
translation(d,D̂) :–
    translation(a,â),                 ←
    translation(B,B̂),
    analogy(a,B,C,d),
    translation(C,Ĉ),
    analogy(â,B̂,Ĉ,D̂),
    assert(translation(d,D̂)).
```

% translation program proper

translation(d,$\widehat{D}$) :–
  translation(a,$\widehat{a}$),
  translation(b,$\widehat{b}$), $\leftarrow$
  analogy(a,b,C,d),
  translation(C,$\widehat{C}$),
  analogy($\widehat{a}$,$\widehat{b}$,$\widehat{C}$,$\widehat{D}$),
  assert(translation(d,$\widehat{D}$)).

% translation program proper

```
translation(d,D̂) :-
    translation(a,â),
    translation(b,b̂),
    analogy(a,b,C,d),          ←
    translation(C,Ĉ),
    analogy(â,b̂,Ĉ,D̂),
    assert(translation(d,D̂)).
```

% translation program proper

$\texttt{translation}(d,\widehat{D})$:–
    $\texttt{translation}(a,\widehat{a})$,
    $\texttt{translation}(b,\widehat{b})$,
    $\texttt{analogy}(a,b,c,d)$,
    $\texttt{translation}(c,\widehat{C})$, $\leftarrow$
    $\texttt{analogy}(\widehat{a},\widehat{b},\widehat{C},\widehat{D})$,
    $\texttt{assert}(\texttt{translation}(d,\widehat{D}))$.

% translation program proper

translation(d,$\widehat{D}$) :–
    translation(a,$\widehat{a}$),
    translation(b,$\widehat{b}$),
    analogy(a,b,c,d),
    translation(c,$\widehat{c}$), $\leftarrow$
    analogy($\widehat{a}$,$\widehat{b}$,$\widehat{c}$,$\widehat{D}$),
    assert(translation(d,$\widehat{D}$)).

% translation program proper

translation(d,$\widehat{D}$) :–
    translation(a,$\widehat{a}$),
    translation(b,$\widehat{b}$),
    analogy(a,b,c,d),
    translation(c,$\widehat{c}$),
    analogy($\widehat{a}$,$\widehat{b}$,$\widehat{c}$,$\widehat{D}$), $\leftarrow$
    assert(translation(d,$\widehat{D}$)).

% translation program proper

translation(d,$\widehat{d}$) :–
    translation(a,$\widehat{a}$),
    translation(b,$\widehat{b}$),
    analogy(a,b,c,d),
    translation(c,$\widehat{c}$),
    analogy($\widehat{a}$,$\widehat{b}$,$\widehat{c}$,$\widehat{d}$), $\leftarrow$
    assert(translation(d,$\widehat{d}$)).

% translation program proper

```
translation(d,d̂) :-
    translation(a,â),
    translation(b,b̂),
    analogy(a,b,c,d),
    translation(c,ĉ),
    analogy(â,b̂,ĉ,d̂),
    assert(translation(d,d̂)).        ←
```

% translation program proper

$\texttt{translation}(d,\widehat{d})$:–
    $\texttt{translation}(a,\widehat{a})$,
    $\texttt{translation}(b,\widehat{b})$,
    $\texttt{analogy}(a,b,c,d)$,
    $\texttt{translation}(c,\widehat{c})$,
    $\texttt{analogy}(\widehat{a},\widehat{b},\widehat{c},\widehat{d})$,
    $\texttt{assert}(\texttt{translation}(d,\widehat{d}))$.

% database of facts (bicorpus)

```
translation(s1,ŝ1).
translation(s2,ŝ2).
        ⋮
translation(sn,ŝn).
translation(d,d̂).
```

$\texttt{translation}(s_1,\widehat{s_1})\,.$

$\texttt{translation}(s_2,\widehat{s_2})\,.$

$\vdots$

$\texttt{translation}(s_n,\widehat{s_n})\,.$

$\texttt{translation}(d,\widehat{d})\,.$

Translation examples

このツアーの $\boxed{\text{料金}}$ はいくらですか。

/kono tuā no $\boxed{\text{ryōkin}}$ wa ikura desu ka/

'this tour GEN $\boxed{\text{fee/price (N)}}$ TOPIC how-much to-be INTERR'

271	*How much does this tour $\boxed{\textit{cost}}$?*
160	*How much do you $\boxed{\textit{charge}}$ for this tour?*
141	*What's the $\boxed{\textit{price}}$ of this tour?*
94	*What does this tour $\boxed{\textit{cost}}$?*
43	*What's the $\boxed{\textit{price}}$ of the tour?*
6	*What is the $\boxed{\textit{price}}$ of the tour?*
6	*How much is the green fee?*

Translation examples

胃が $\boxed{\text{痛い}}$ んです。
/i ga $\boxed{\text{itai}}$ n desu/
‘stomach NOM $\boxed{\text{painful (A)}}$ INSIST to-be’

1744	*I have a stomach $\boxed{\textit{ache}}$.*
552	*My stomach $\boxed{\textit{hurts}}$*
124	*I’ve got a stomach $\boxed{\textit{ache}}$.*
56	*Do you have a stomach ache.*
51	*Do you have a stomach ache?*
50	*I have a stomach ache?*
2	*My stomach $\boxed{\textit{hurts}}$ me.*
1	*I have an abdominal $\boxed{\textit{pain}}$ in my stomach.*
1	*I have a $\boxed{\textit{pain}}$ in my stomach.*
1	*I have a soare throat.*

Translation examples

コーヒーのおかわりを頂けますか。
/kōhī no o-kawari wo itadakemasu ka/
'coffee GEN POLITE change/again ACC can-receive INTERR'

2318	*I'd like another cup of coffee.*
2296	*May I have another cup of coffee?*
1993	*Another coffee, please.*
1982	*May I trouble you for another cup of coffee?*
1982	*Can I get some more coffee?*
530	*Another cup of coffee, please.*
516	*Another cup of coffee.*
466	*Can I have another cup of coffee?*
337	*May I get some more coffee?*
205	*May I trouble you for another cup of coffee, please?*

Translation examples

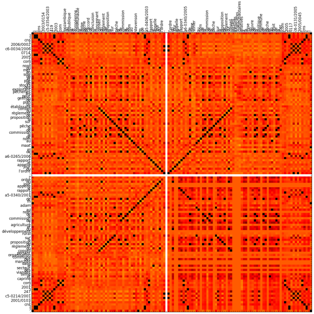
小銭を混ぜてください。
/kozeni wo mazete kudasai/
'coins/small-change ACC to-mix deign/if-you-please'

924	*Can you include some small change?*
922	*Can you include some small change, please?*
899	*Would you include some small change?*
896	*Include some small change, please.*
895	*I'd like to have smaller bills mixed in.*
895	*Please change this into small money.*
895	*Will you include some small change?*
885	*Could you include some small change, please?*
880	*May I have some small change, too?*
877	*Please give me some small change as well.*

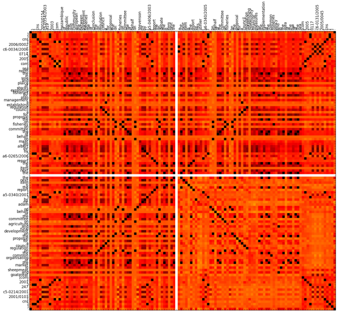
Next research direction

- Machine translation by analogy
- using word vector representations
- and sub-sentential alignment matrices
- to translate piece by piece.

French



English



Contents

Waseda University, IPS, EBMT/NLP Lab

Merci de votre attention.
ご静聴有難うございました。