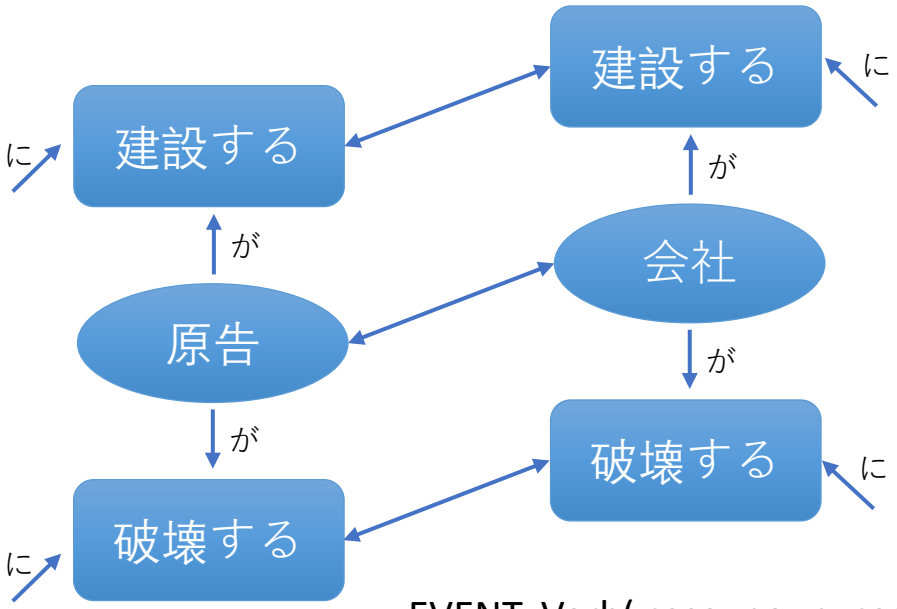
A Study on Descriptive Patterns Based on Similarity Classes of Individual Constants

個体の類似クラスに基づく記述的類似性

Knowledge Base Lab.
M2
Ruipeng Wang

Descriptive similarity

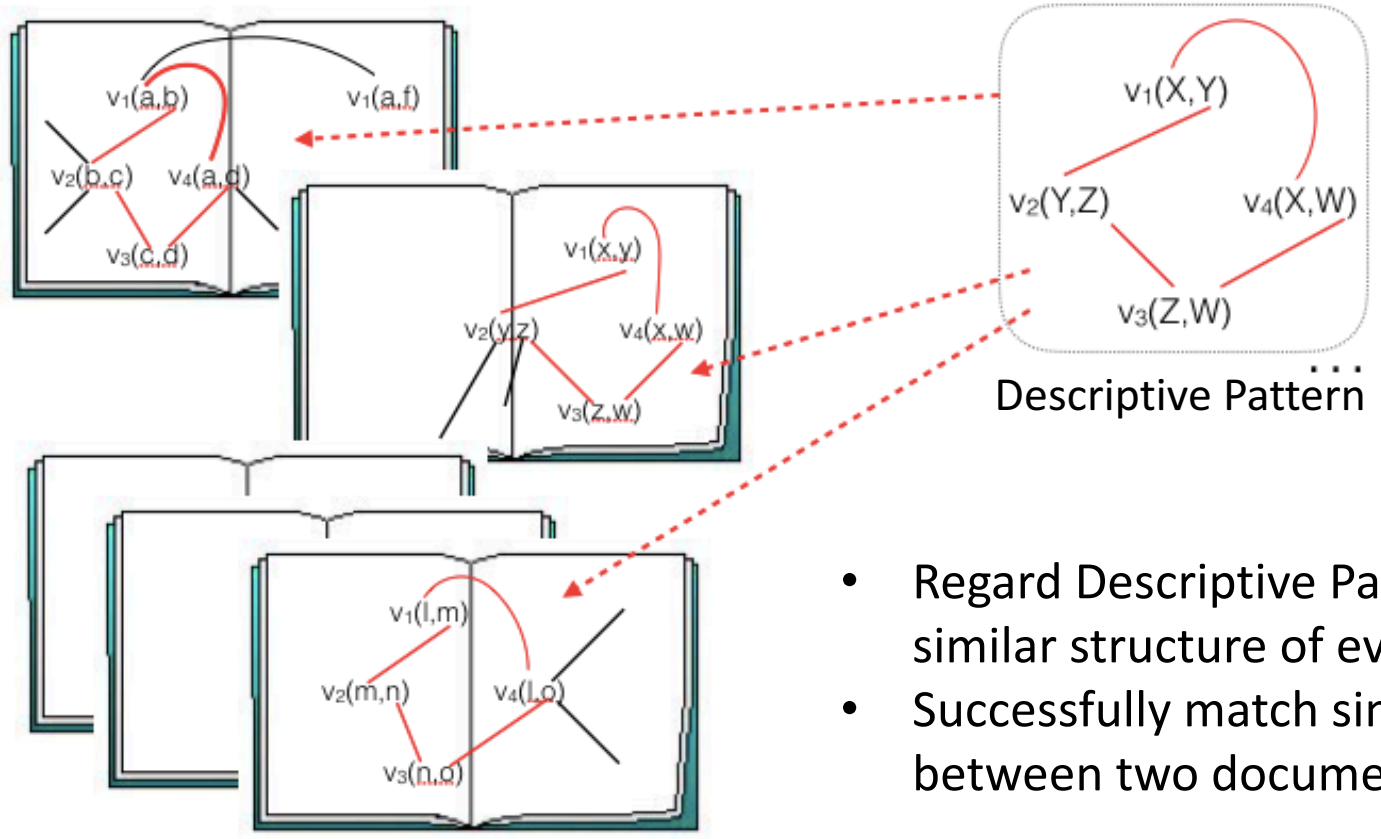


EVENT: Verb(case: noun, case: noun,…)
ex.建設する(が/原告,に/…)

One to one correspondence ：
Nouns come with same verb-cases from different sentences

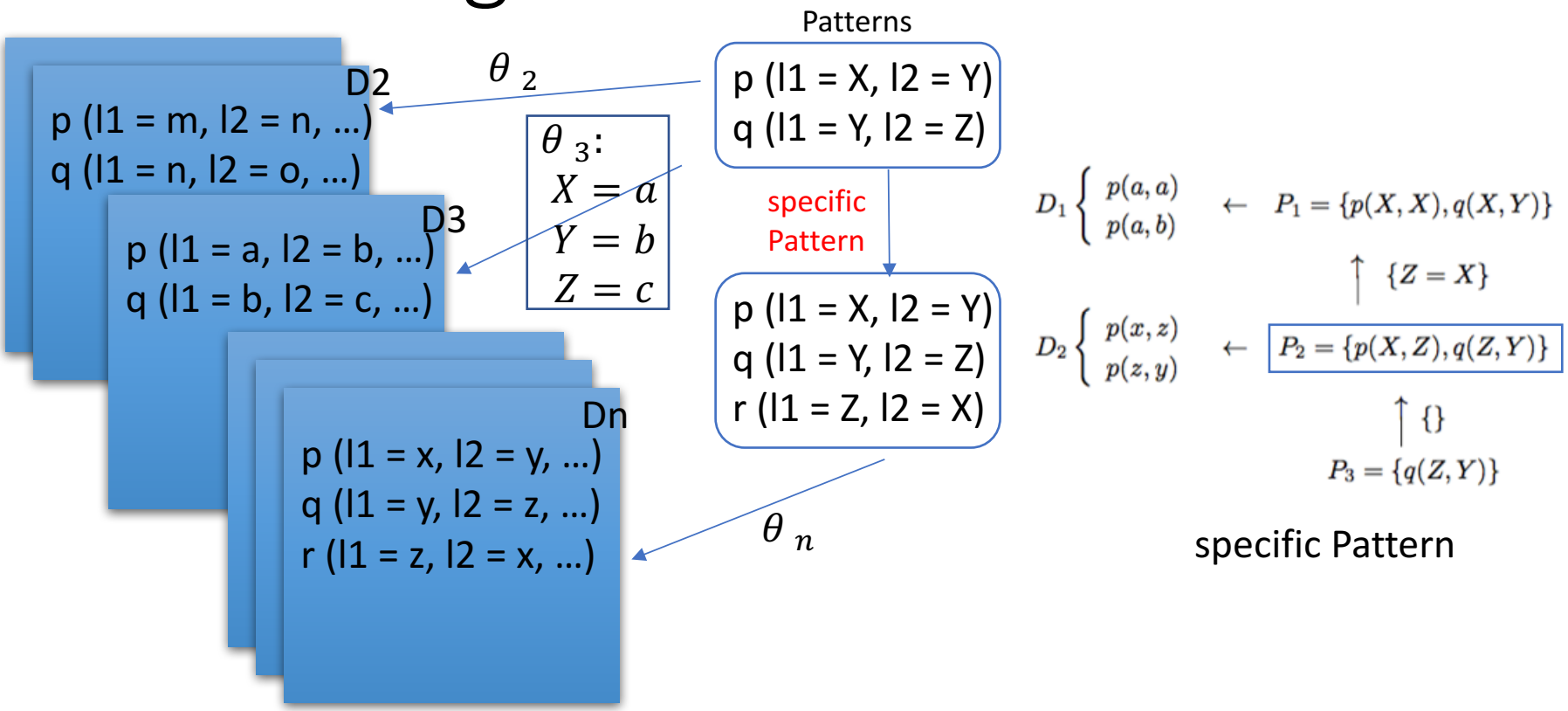
Previous research

X. Zhang, Feb. 2016, Master Thesis



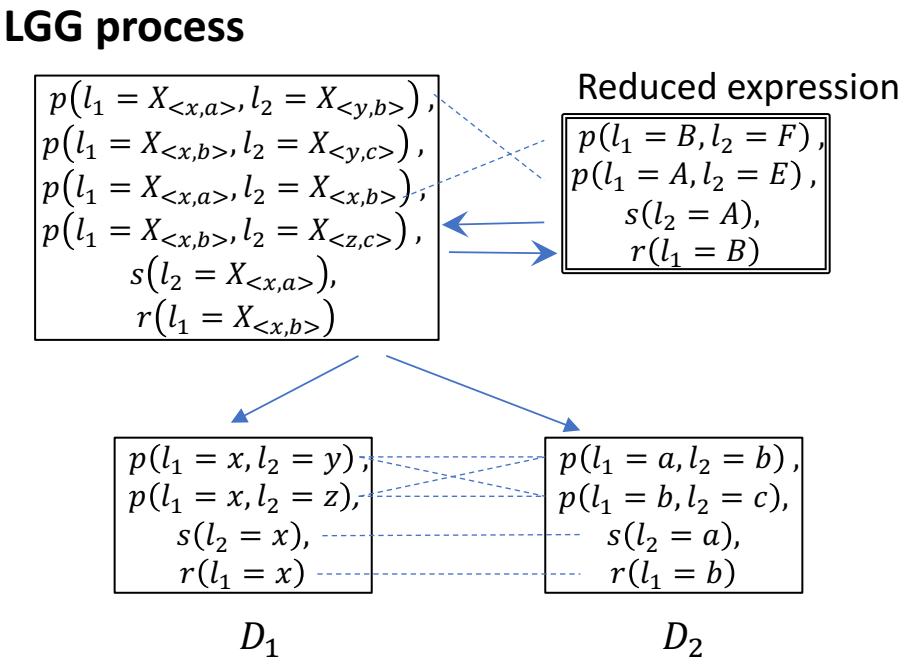
- Regard Descriptive Pattern as a similar structure of event graphs.
- Successfully match similar structure between two documents

Research target



- Regard group of event as patterns, and extract most specific frequency pattern as descriptive pattern.
- Consider multi-document situation, extract patterns supported by multi-document. (a kind of Data Mining instead of matching.)
- Phase 1: extract similarity class of Individual Constants.(Input of next phase)
- Phase 2: Re-construct descriptive patterns. (Depend on the result of Phase 1)

Least General Generalization



- Time complexity will be very large will computing LGG.
- We propose to use neither pairs nor tuples: Similarity classes of individuals over domains: frequent closures (intent of formal concepts)

Maximal closures

role set with maximal roles is complete corresponding with DP

Descriptive Pattern, P:

$\{V_0(l_1 = X, l_2 = Y, \ldots),$
$V_1(l_1 = Y, l_2 = X, ,), \ldots\}$

verb-role pairs played by the variable X

$role_P(X) = \{v_0(l_1),\ v_1(l_2)\}$

$$\begin{cases} maximal: \\ \{v_0(l_1),\ v_1(l_2)\} \\ non-maximal: \\ \{v_0(l_1)\}, \{v_1(l_2)\} \end{cases}$$

$a \ \ldots$	$x \quad y$	$\cdot m \cdot$
$const(D_1)$	$const(D_2)$	$\ldots$

$D_1 \qquad D_2 \qquad D_3$

Inner Predicate composition:

recover the co-occurrence information of variables.

$v_0(l_1 = X), v_2(l_2 = Y)$
$v_1(l_1 = Y), v_2(l_2 = X)$
....

Pattern whose event has just one role description

- descriptive pattern can be re-constructed from maximal closure

Inner Predicate composition

Any variable in DP has a closure: $A = \{v_0(l_1), v_1(l_1)\}$

Conversely we make primitive patterns from them, and
compose the primitive patterns to get predicates with more arguments

Primitive pattern for closures

$pp(A) = v_0(l_1 = A), v_1(l_1 = A)$

$v_1(l_2 = B), v_2(l_1 = B)$

$v_2(l_2 = C)$

$v_3(l_1 = D)$

$v_4(l_1 = E), v_3(l_2 = E)$

Inner predicate composition

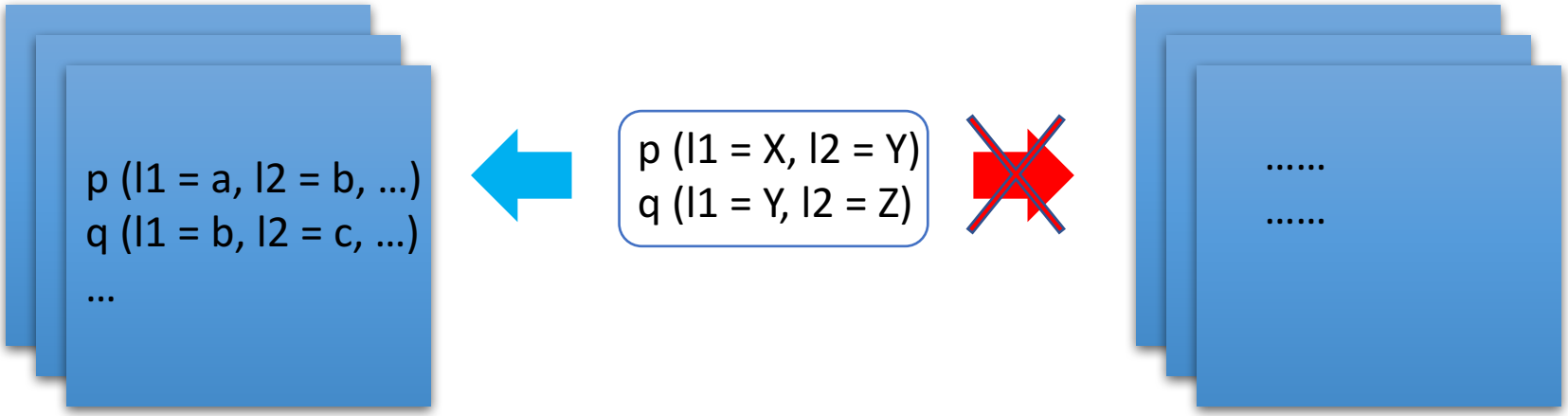
$v_0(l_1 = A), v_1(l_1 = A, l_2 = B),$
$v_2(l_1 = B, l_2 = C)$

$v_4(l_1 = E), v_3(l_2 = E, l_1 = D)$

More specific pattern towards DP

- Beam Search Algorithm to extract some main descriptive patterns.

Requirement: Minimal Support

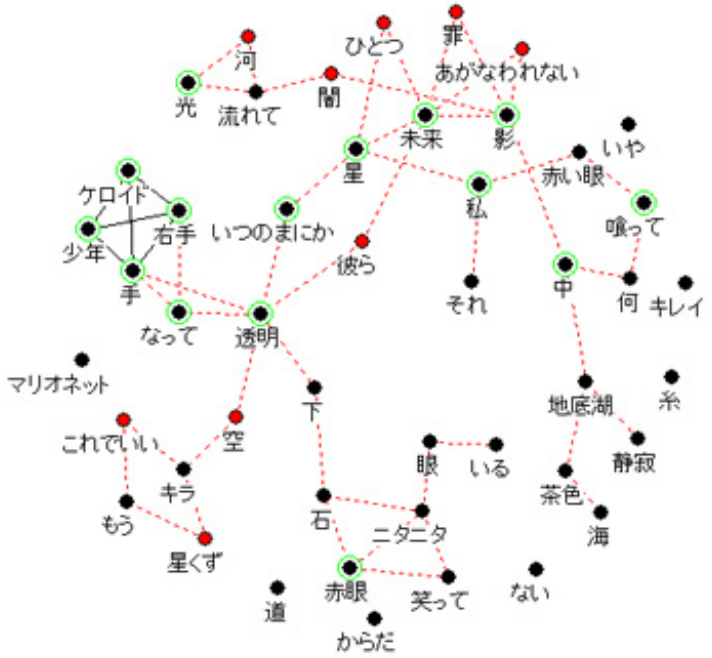


- a pattern is supported by multi-document.
- We set τ donates the percentage of document that support a pattern.
- ex. For a Pattern P , $\tau = 1$ means all documents support P.

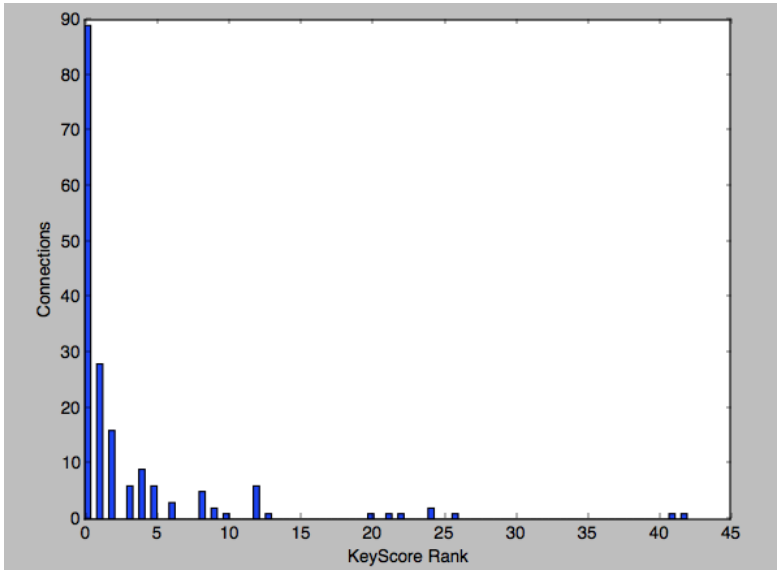
- Our pattern is required to be supported by multi-document.

Requirement: KeyGraph

- Closures to be candidate include at least one High KeyScore noun.
- KeyScore reflects the importance of a noun in a document.
- Feasibility: connections between different events mainly generated by high KeyScore nouns.



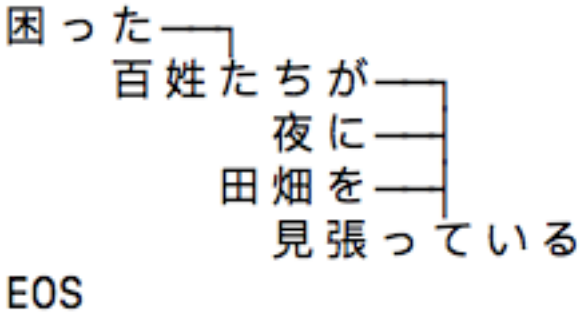
black: high frequency words（hf）
red：not hf but close contact to hf
green： high KeyScore words



KeyScore Rank words over 45th are omitted

Data

- We can use KNP tool to extract events with verb-case information.
  - 困った百姓たちが夜に田畑を見張っている



  - 見張る(百姓/ガ, 夜/ニ,田畑/ヲ)
- We use 100 precedents and extract 84,550 events.
- Average length of the precedents is 45,000 Japanese words.

Experiment: Stories

data: two short Japanese stories, t = 1.

part of closures extracted

* [荒らす/ヲ] [町001, 八百屋001, 田畑002]
* [持つ/ヲ] [光001, 火縄銃002]
* [困る/カ] [若者001, 村人002]
* [現れる/カ] [化け物001, 鹿002, 牛001, 老人001]

original story 1

…ところがある夜から、金色の二つの光を持った化け物が現れ、町の八百屋を荒らして回るようになった。 困った若者たちが、夜の八百屋を見張っていると…

original story 2

…夜な夜な2頭のつがいの大鹿が現れ、田畑を荒らしまわるので、村人は大層困っていた…

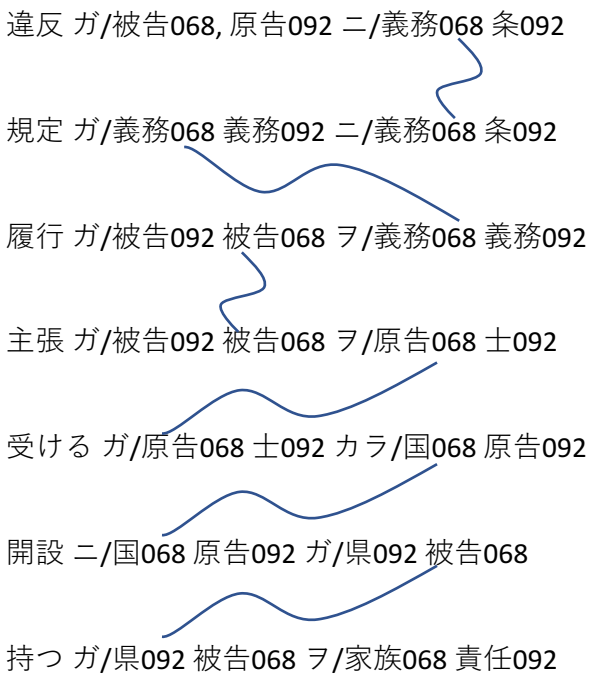
あやしい牛	あばれ鹿
町	田畑
若者	村人
光	火縄銃
牛	鹿

Experiment: Precedents

data: three precedents include two similar ones and a very different one, t = 2/3

words(noun)	Event(s)	KeyGraph(s)	MFC(s)
76083	0.15	10.31	8.27

105 maximal closures founded by these three document, and 98 closures are supported by the similar ones. Over 93%.



part of descriptive pattern re-constructed by extracted maximal closures

Summary and Future Work

- Give the exact definition of Descriptive Pattern.
- Validate the feasibility of KeyGraph.
- Extract similarity classes from multi-document.

- Realize the re-construct descriptive pattern from maximal closures.
- Adjust the output of KNP, fix the results.