Japanese Predicate Argument Structure Analysis Exploiting Argument Position and Type

Graduate School of Information Science Nara Institute of Science and Technology (NAIST)

Yuta Hayashibe, Mamoru Komachi, Yuji Matsumoto

The 5th International Joint Conference on Natural Language Processing (IJCNLP)

2011.Nov.9 14:20-14:45 @Chaing Mai

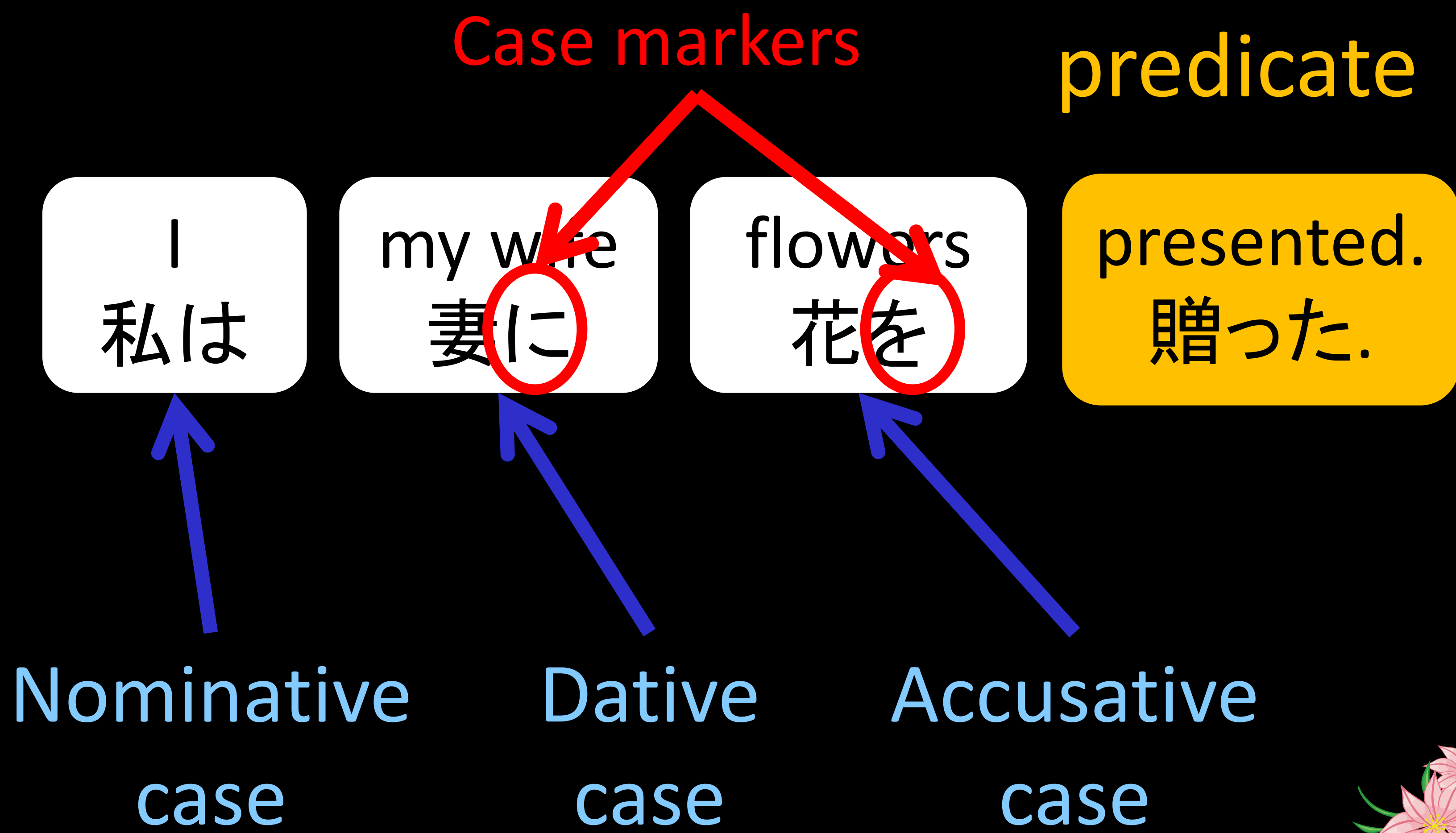
Predicate-Argument Structure Analysis (PASA)

to identify the argument for each case

Case = semantic relation between phrases

In short, each case corresponds to …

Case	Semantic Role
Nominative	Agent
Accusative	Direct object
Dative	Indirect object



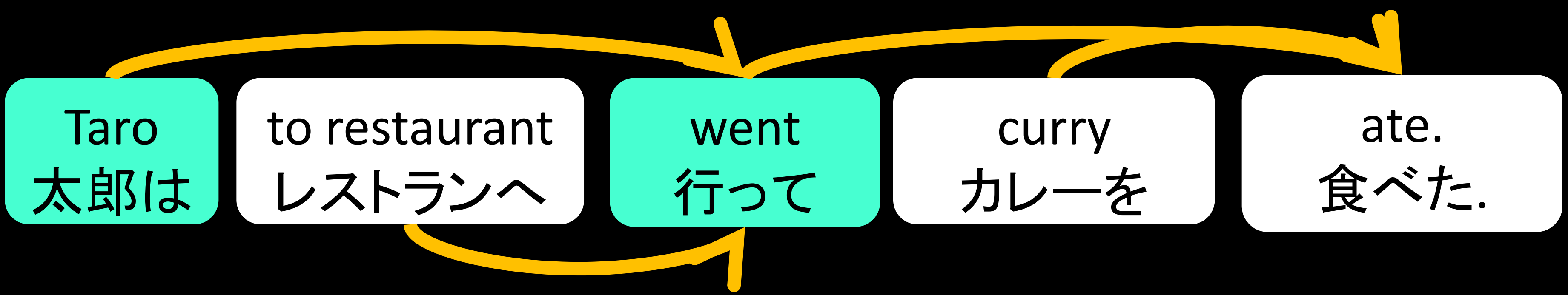
Useful for Machine Translation, Information Extraction, summarization and so on.

Arguments are classified into…

3 types

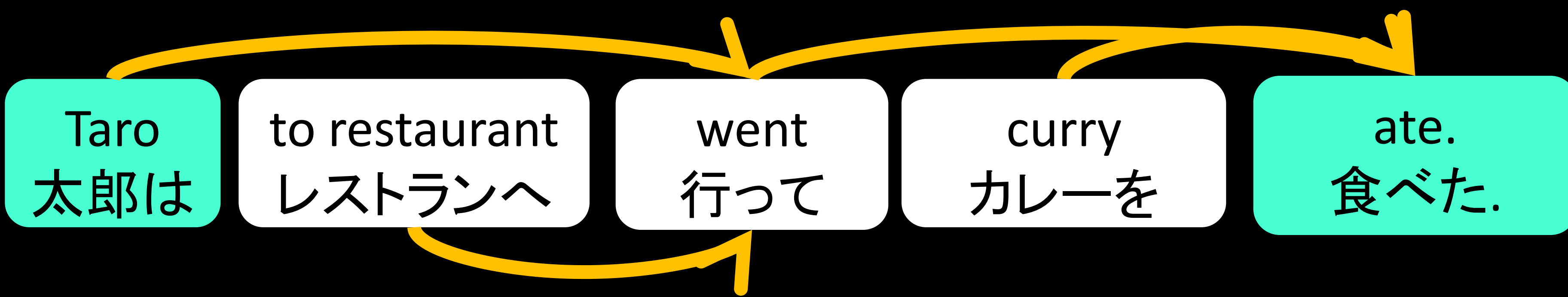
according to their positions relative to the predicates

INTRA_D



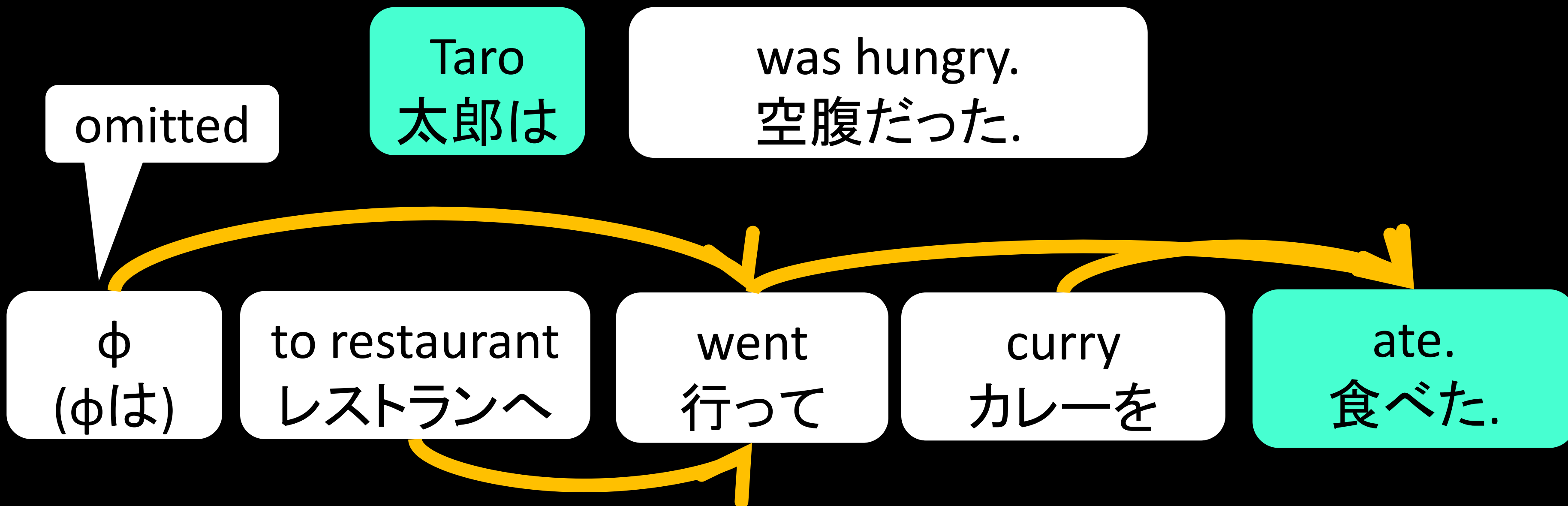
intra-sentential arguments
that have direct syntactic dependency with the predicates

INTRA_Z



intra-sentential arguments
 that doesn’t have direct syntactic dependency with the predicates

INTER



inter-sentential arguments
(which are not in the same sentence)

3 argument types

INTRA_D

INTRA_Z

INTER

The difficulty of analysis

Comparatively easy

	Nominative	Accusative	Dative
INTRA_D	75.6	88.2	89.5
INTRA_Z	30.2	11.4	3.7
INTER	23.45	9.32	11.76

Difficult

F-value of PASA [Taira08]

Why is it more difficult to identify the arguments of INTRA_Z and INTER than INTRA_D?

	function word , directly dependency…	Arguments (e.g: nominative)
INTRA_D	✓	57.1%
INTRA_Z	✗	30.5%
INTER	✗	12.4%

Poverty of features

limited amount of training examples

2 proposals

Proposal1:

a novel feature
with an improved similarity measure
between argument positions
of two predicates
that take semantically similar arguments

Argument position

the position where the argument of a predicate may occur

What is the argument of nominative case of “surrender”?

police *wa*-particle hanako *wo*-particle arrested
警察 は 花子 を 逮捕した.

Police arrested Hanako.

I had surrendered that heard
私は (ϕ が) 自首した と 聞いた.

I heard that ϕ *had surrendered.*

Arrest
- nominative case : Police:
- accusative case : Hanako

Surrender
- nominative case : ?

Need the knowledge that an an rested person is more likely to be a person who has surrendered than an arrestee

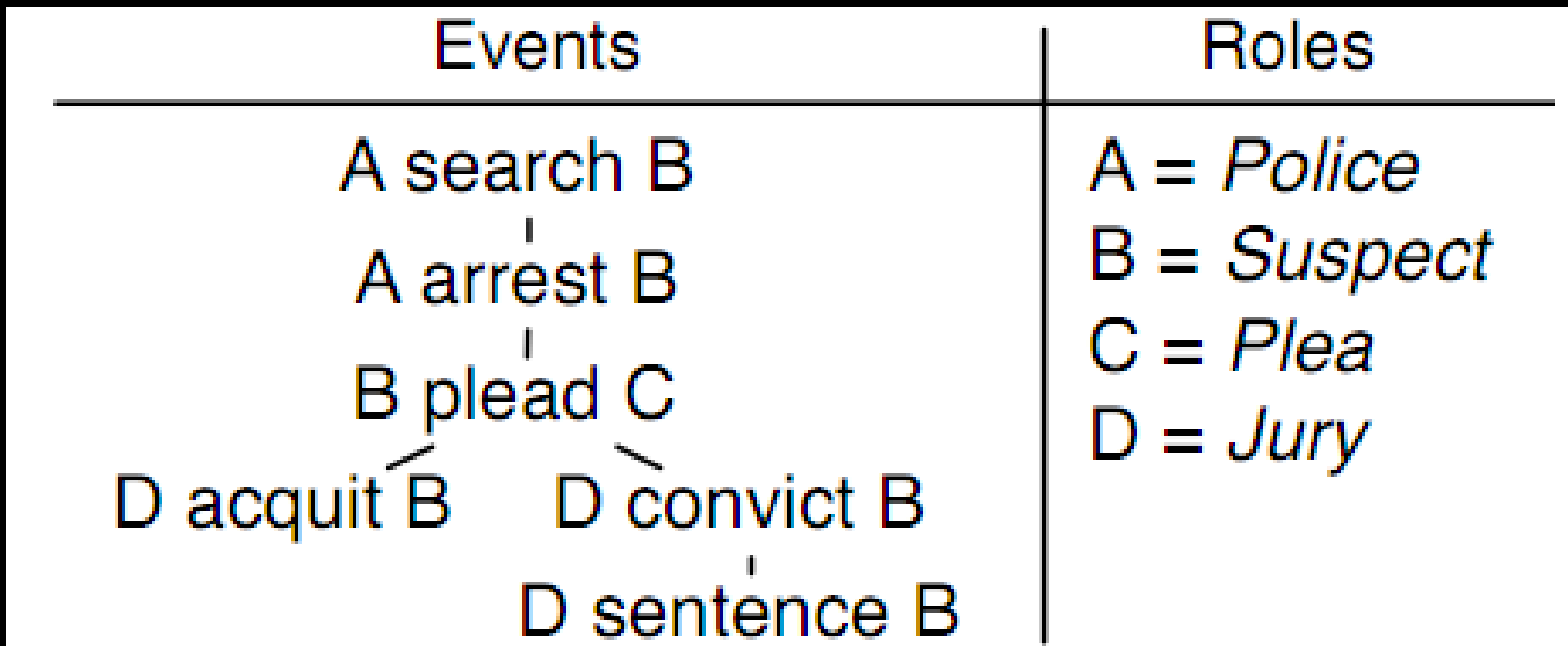
Arrest - accusative case

look alike

Surrender - nominative case

Large amount of Texts

Get the knowledge using co-reference chain?



[Chambers et al.09]

(1)Strong dependency on co-reference chain resolutions

(2)Need of many documents (does not use any non co-referring nouns)

Extract large amount of triples with a dependency parser

Verb + case marker + noun phrase

1599	が 自首する (surrender)	5651	が 逮捕する (arrest)	82112	を 逮捕する (arrest)
136	-者 person	702	署員 PS staff	15285	人 person
117	犯人 criminal	698	警察 police	8484	-者 person
96	彼 he	376	警察官 police officer	5563	男 man
68	人 person	368	-署 police station	2804	-名 name
63	男 man	230	県警 prefectural police	1188	犯人 criminal
36	-犯 crime	177	-員 -staff	1185	男性 male
30	少年 boy	153	府警 prefectural police	763	-ら -s
26	高校生 high-scool student	137	当局 authorities	671	おまえ you

が(ga):nominativecase を(wo):accusative case

Extracted from web corpus about 5milion sentences[Kawahara et al.09]

Similarities between argument positions calculated with WEB corpus

Sim(A,B)☰1- JenshenShanon(A,B)

	nominative case of surrender	nominative case of arrest	accusative case of arrest
nominative case of surrender	-	0.4590	0.7568
nominative case of arrest		-	0.4861
accusative case of arrest			-

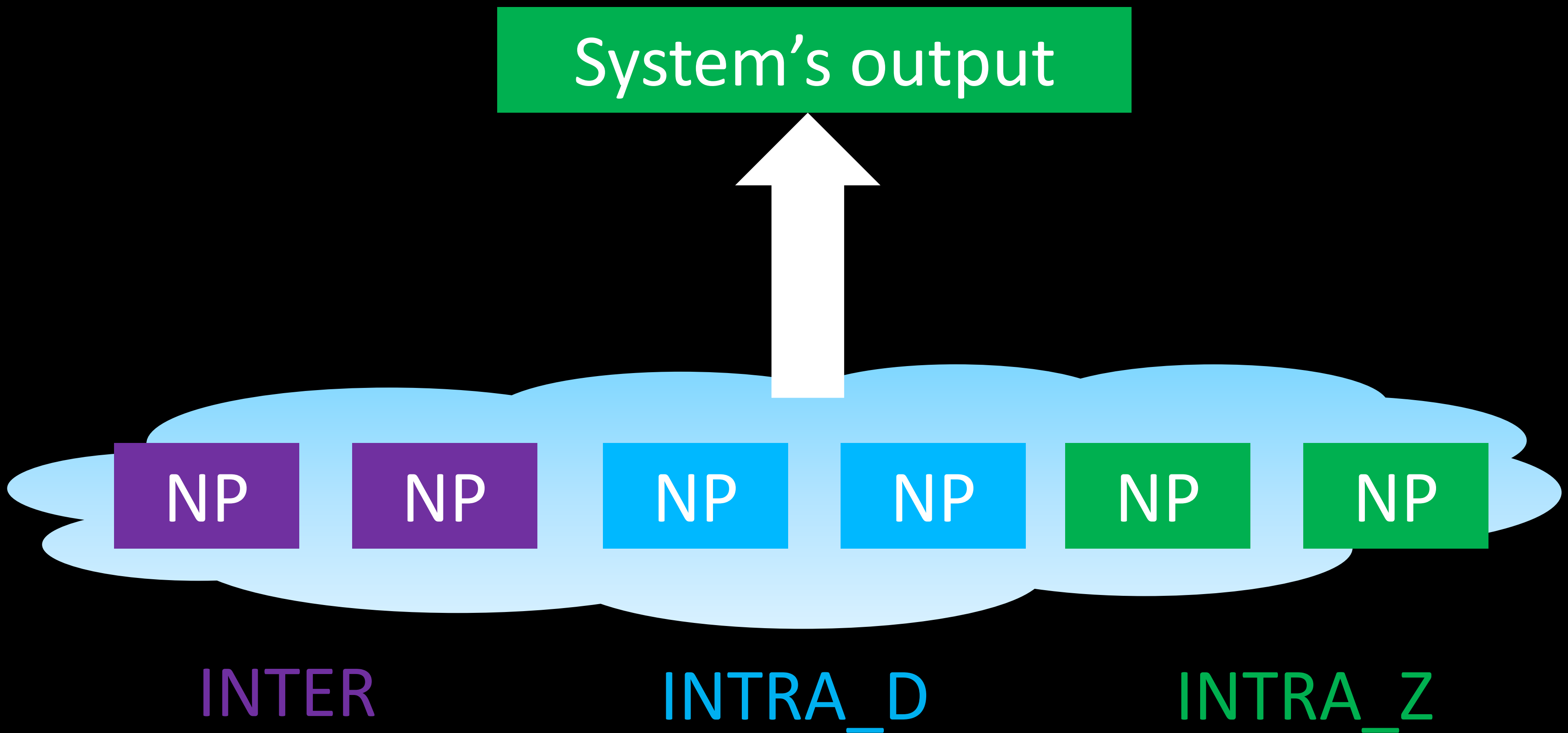
Argument position …the position where the argument of a predicate may occur.

Advantages of our proposed method

- In general, the accuracy of dependency parsing is higher than that of co-reference chain resolution.
- This method can use all triples in documents.

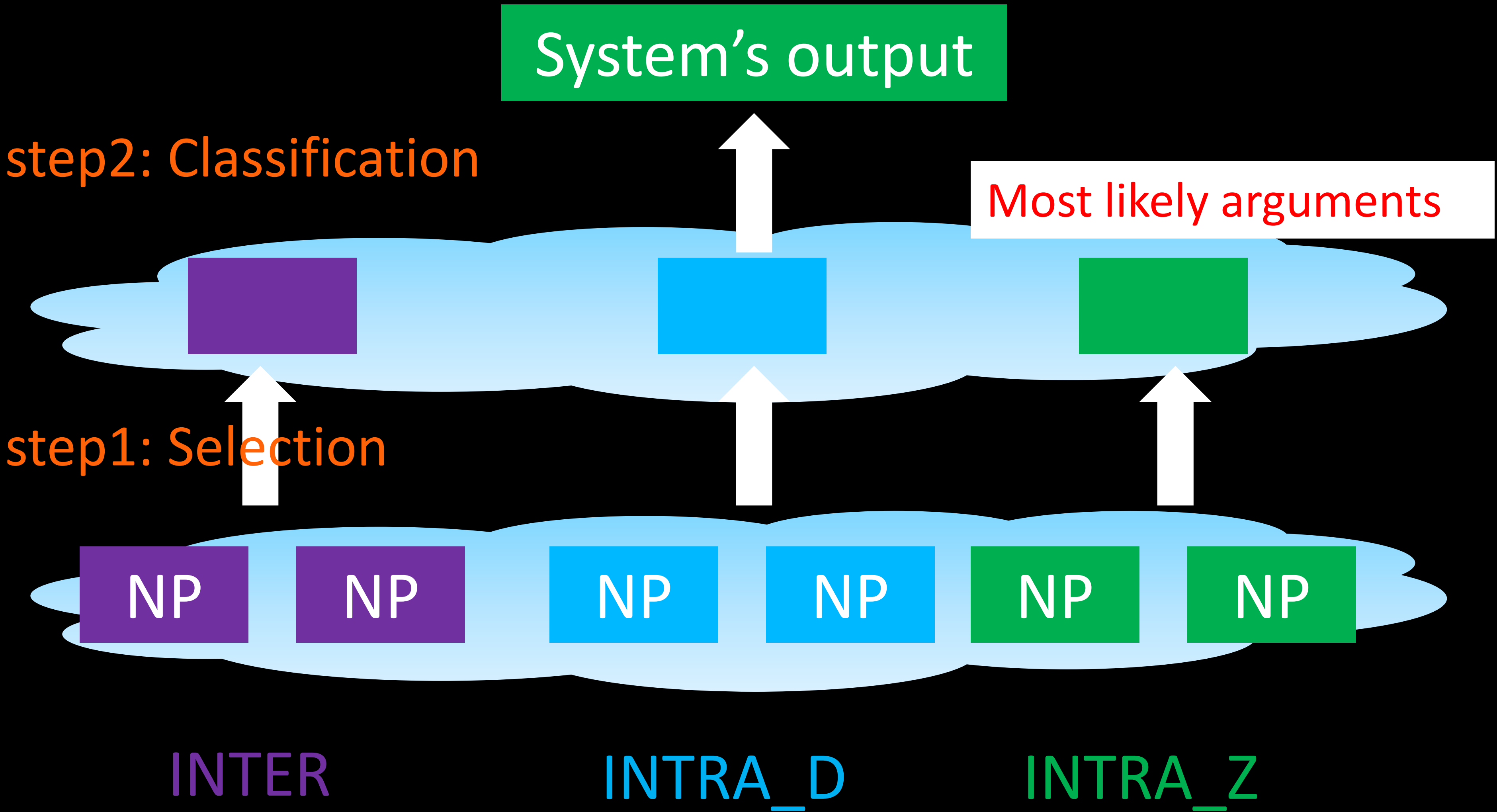
Proposal2:
Selection-and-Classification Model Considering Argument Types

With a single model in past work

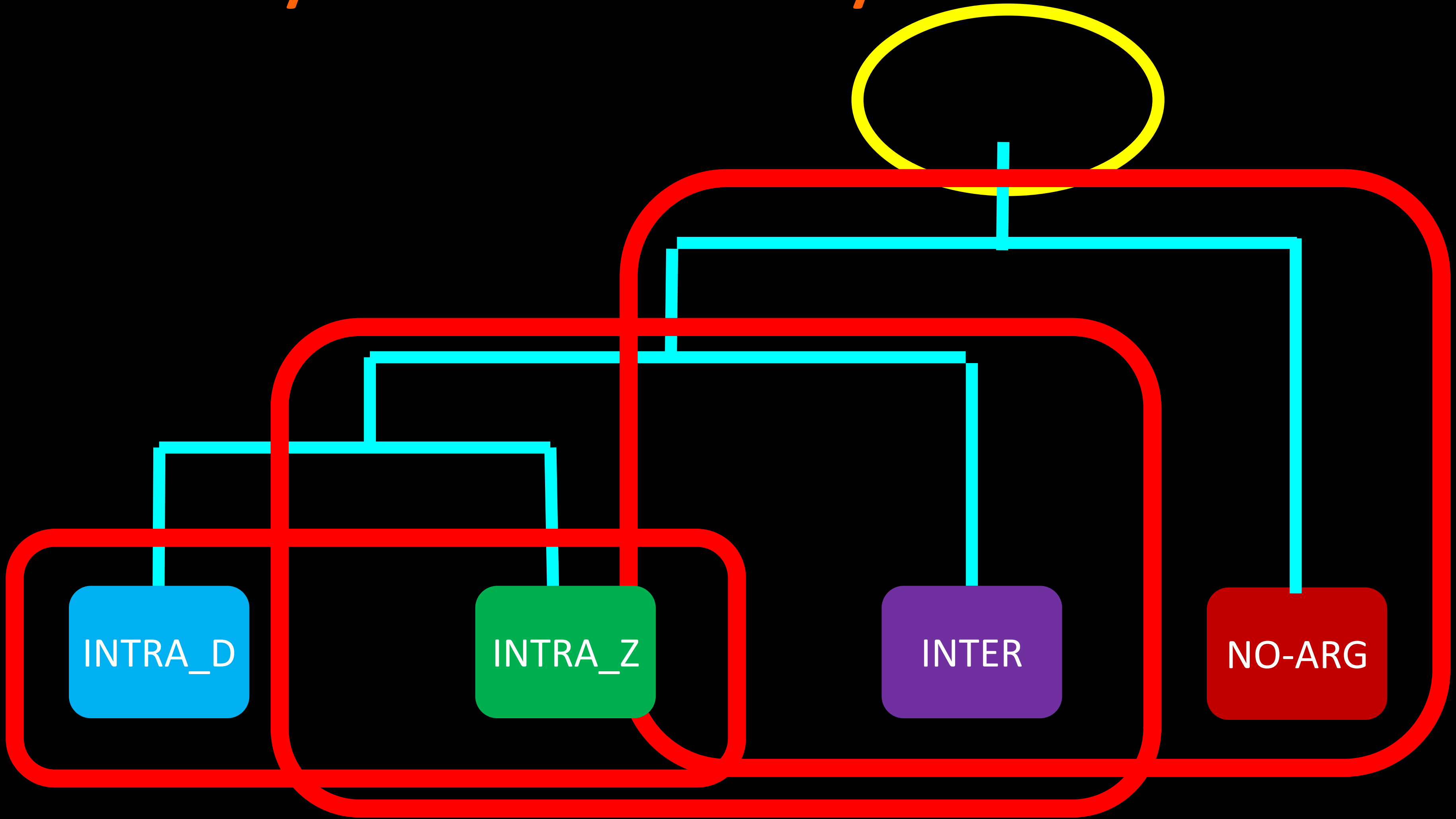


Difficulty in distinguishing the marginal cases where two candidates have different argument types

Selection-and-classification model



Classify with 3 binary clasifications



Experiments

Experiment settings

Case	Nominative case
Clasifier	Support Vector Machine(linear kernel)
Common features	Features proposed by [Iida et al.07]
Most likely argument selection	Tournament model[Iida et al.10]
Dataset	NAIST text corpus1.4β (2917 articles) source: Japanese newswire texts
Test method	5-fold cross validation
Assumption	The results of co-reference resolution and former PASA are correct.

Result

We investigate from 3 standpoints

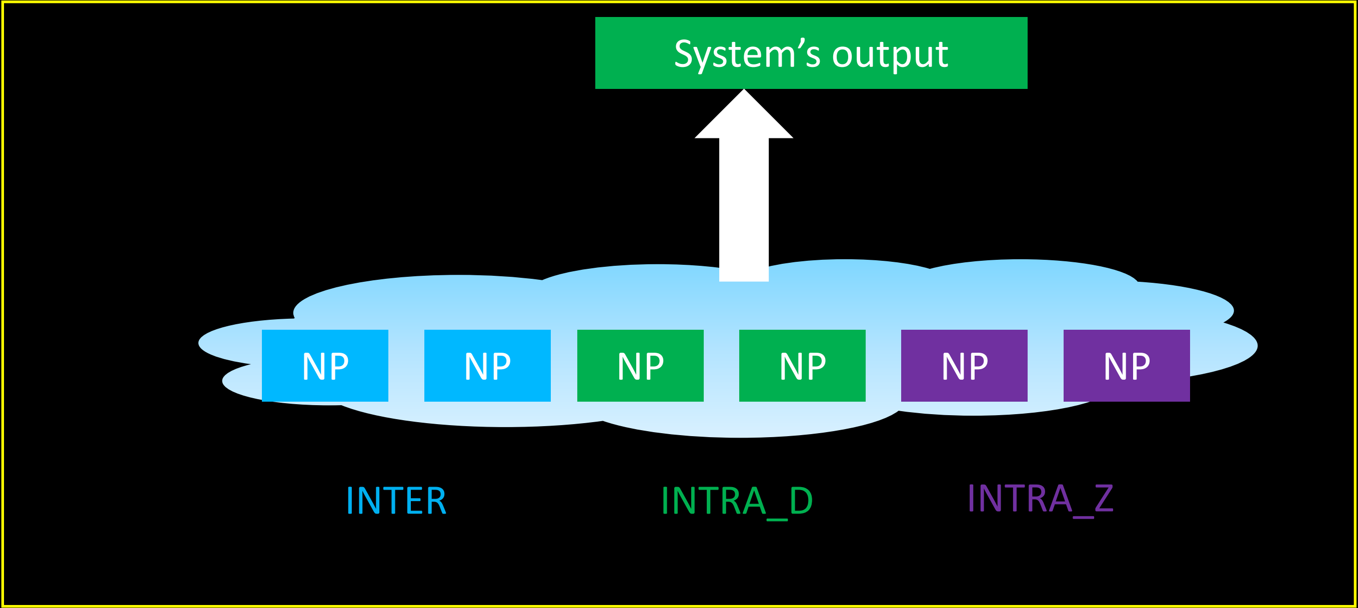
(1) Effect of our proposed S&C model

(2) Effect of our proposed similarity based similarity feature

(3) Effect of In and Out-of-domain

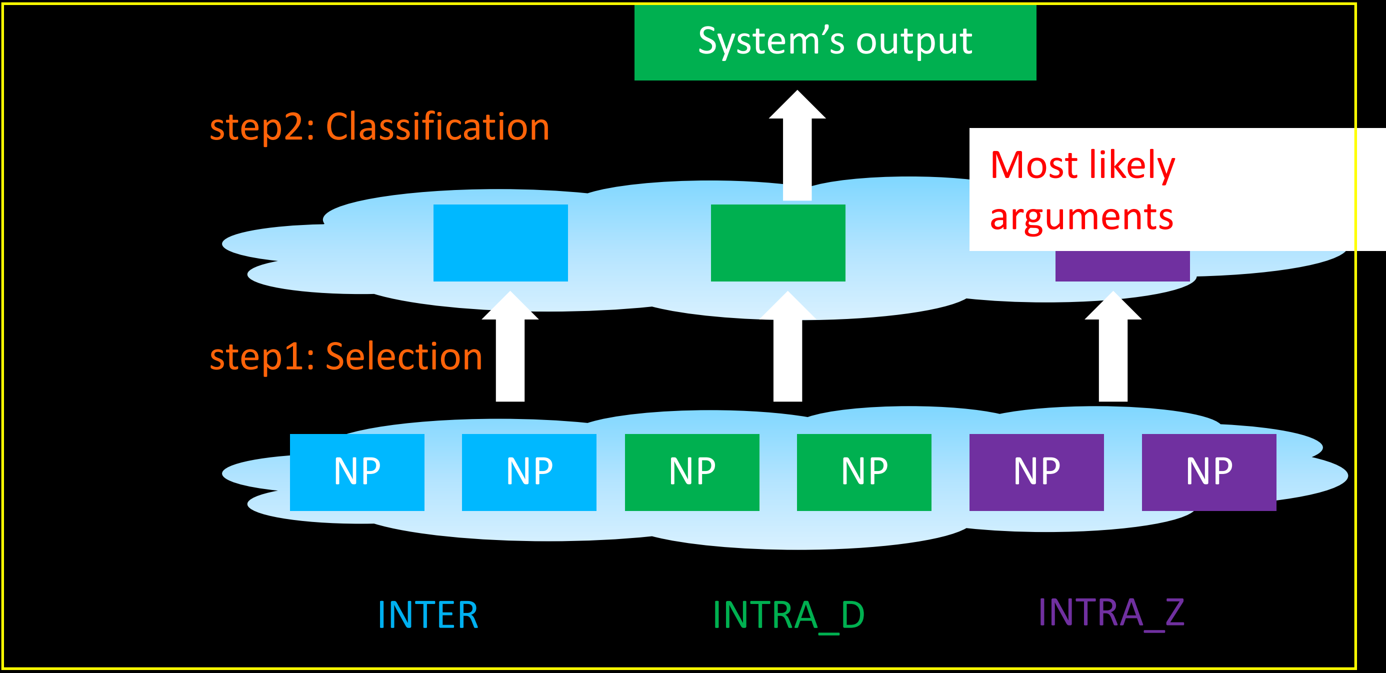
Baseline

Not Considering argument types



S&C model

Considering argument types



Our proposed model is superior to others

	INTRA_D			INTRA_Z			INTER			ALL
	P	R	F	P	R	F	P	R	F	F
Baseline model	80.51	56.86	66.63	27.97	54.91	37.06	19.05	17.70	16.60	50.72
Proposed S&C model	80.71	85.35	82.96	47.57	75.74	46.64	23.79	15.93	19.07	**67.46**
[Taira et al. 08]	-	-	75.53	-	-	30.15	-	-	23.45	57.40

All use standard features

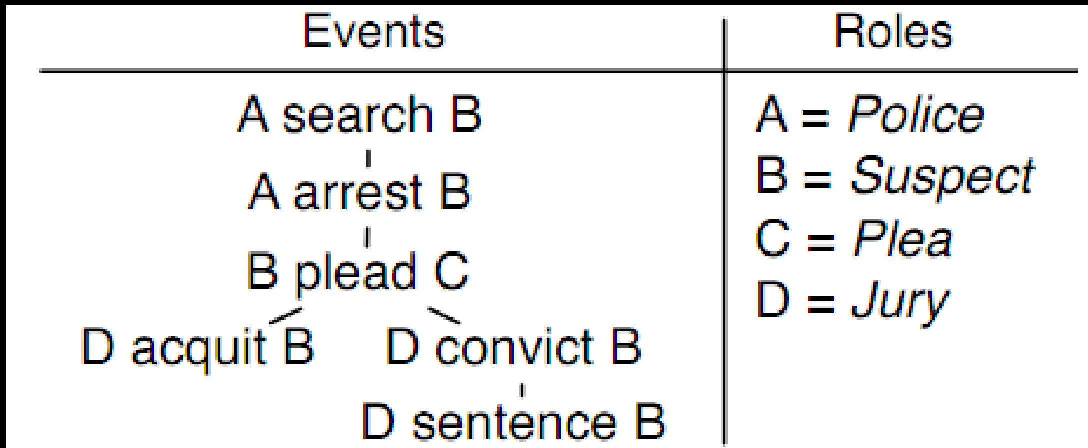
We investigate from 3 standpoints

(1) Effect of our proposed S&C model

(2) Effect of our proposed similarity based similarity feature

(3) Effect of In and Out-of-domain

Co-reference based feature



[Chambers et al.09]

similarity based feature

1599	が(ga) 自首する (surrender)	5651	が(ga) 逮捕する (arrest)	82112	を(wo) 逮捕する (arrest)
136	-者 person	702	署員 PS staff	15285	人 person
117	犯人 criminal	698	警察 police	8484	-者 person
96	彼 he	376	警察官 police officer	5563	男 man
68	人 person	368	-署 police station	2804	-名 name
63	男 man	230	県警 prefectural police	1188	犯人 criminal
36	-犯 crime	177	-員 -staff	1185	男性 male
30	少年 boy	153	府警 prefectural police	763	-ら -s
26	高校生 high-scool student	137	当局 authorities	671	おまえ you

Our proposed similarity based feature is effective

	INTRA_D			INTRA_Z			INTER			ALL
	P	R	F	P	R	F	P	R	F	F
standart features	80.71	85.35	82.96	47.57	75.74	46.64	23.79	15.93	19.07	67.46
+COREF	86.82	88.90	87.85	54.07	52.89	53.47	25.83	20.08	22.58	71.99
+SIM	88.42	91.10	89.74	59.05	58.12	58.58	24.81	19.91	22.08	**74.17**

All use S&C model

COREF : co-reference based feature calculated from Web Texts
SIM: similarity based feature calculated from Web Texts

We investigate from 3 standpoints

(1) Effect of our proposed S&C model

(2) Effect of our proposed similarity based similarity feature

(3) Effect of In and Out-of-domain

Use these sources to calculate similarity between events

Web text

Newswire text

Our proposed feature is effective

	INTRA_D			INTRA_Z			INTER			ALL
	P	R	F	P	R	F	P	R	F	F
standart features	80.71	85.35	82.96	47.57	75.74	46.64	23.79	15.93	19.07	67.46
+SIM_WEB	88.42	91.10	89.74	59.05	58.12	58.58	24.81	19.91	22.08	74.17
+SIM_NEWS	87.00	90.44	88.69	64.76	60.27	62.43	25.63	21.32	23.27	74.45
+SIM_WEB &SIM_NEWS	89.70	91.55	90.62	65.08	61.37	63.17	24.86	21.57	23.08	**75.61**

All use S&C model

SIM_WEB : similarity based feature calculated from Web Texts
SIM_NEWS: similarity based feature calculated from Newswire texts

Result

(1) Our proposed model is effective

(2) Similarity based feature is more effective than co-reference based feature

(3) Measures with In-and-Out-of-domain data work complementary

Error Analysis

Predicate Sense Ambiguity

Light Verb Construction

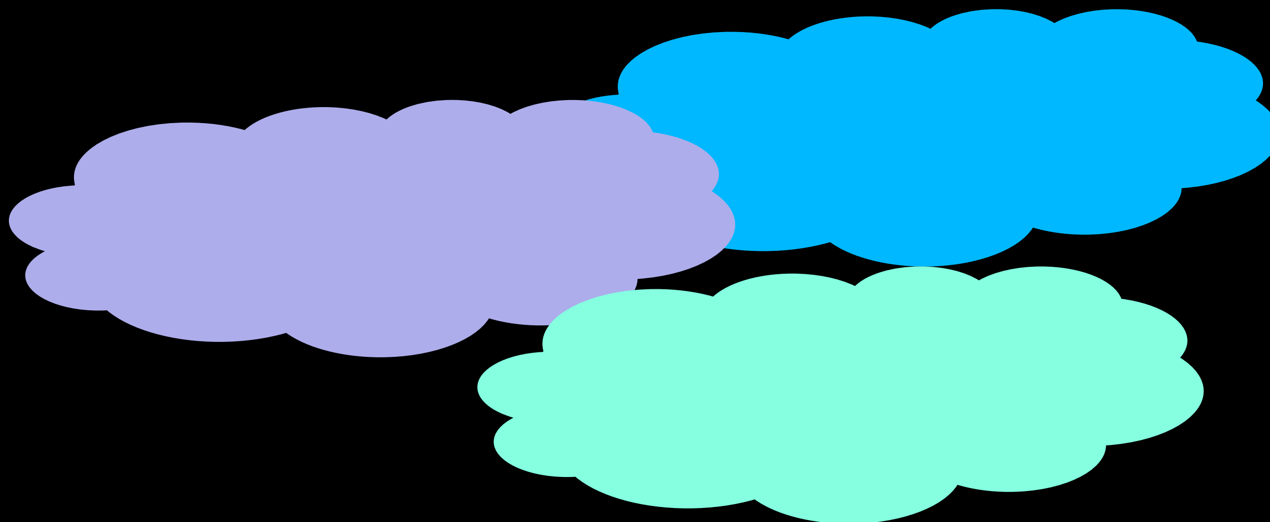
It’s difficult to calculate similarity when the predicate has Sense Ambiguity

（φが）数字を早急に詰める(tsumeru)必要性を強調した。

They emphasized that φ should be brought to an conclusion as soon as possible.

詰める	1. to pack
tsumeru	2. to bring to a conclusion

Ambiguous verbs tend to have a mixture of several distributions of arguments



more difficult when the predicate is more essential verb…

持つ
(have)

取る
(take)

We need WSD

Error Analysis

Predicate Sense Ambiguity

Light Verb Construction

Light Verb Construction

Noun	Particle	Verb
affection	wo	get
影響	を	受ける
[eikyou]	[wo]	[ukeru]

carries the main meaning of this phrase

doesn’t play a central role

Future work

(1) to combine internal argument to take semantic argument into consideration, if the verb is in light verb construction

(2) to perform word sense disambiguation before calculating similarity

(3) to conduct experiments not only on nominative case but also on other cases

Conclusion

Proposal1:

A similarity feature between argument positions using distribution similarity

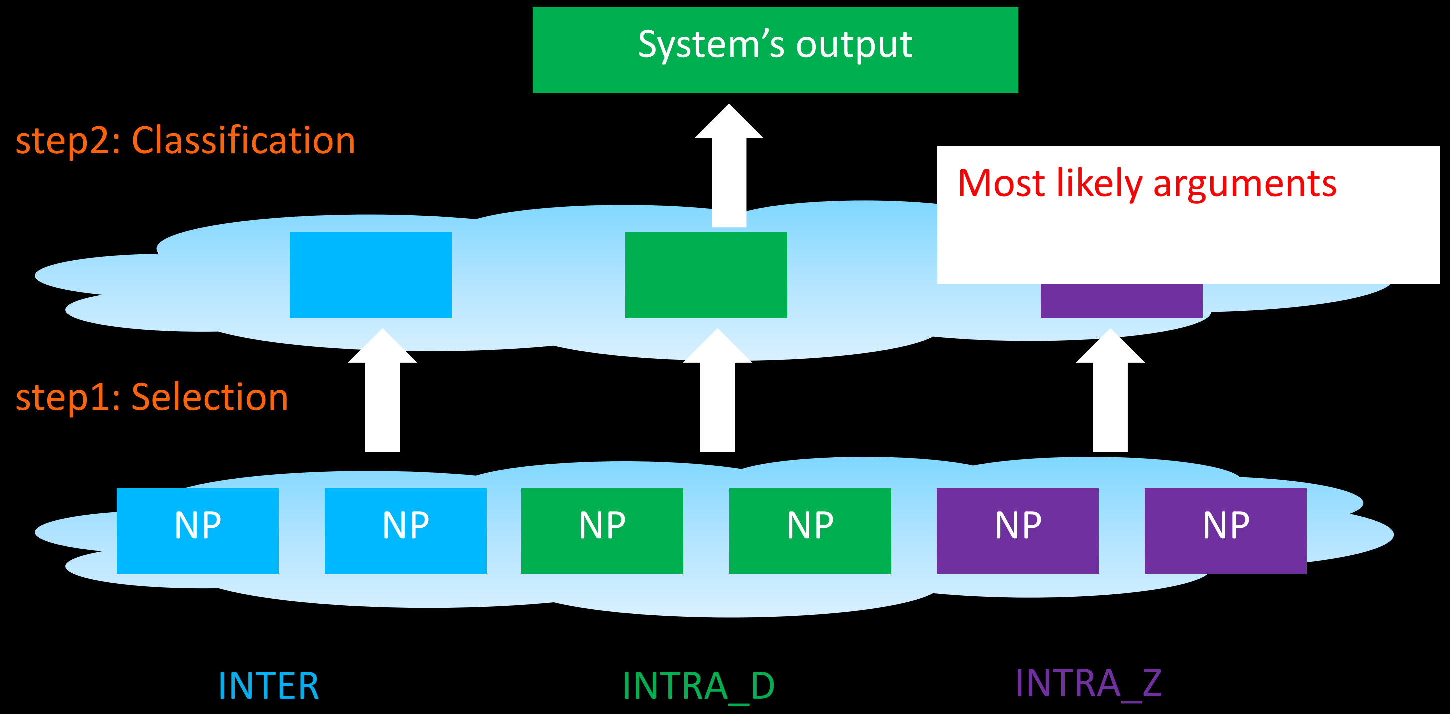
Proposal2:

Selection-and-Classification Model Considering Argument Types

Proposal1:
A similarity feature between argument positions using distribution similarity

1599	が (ga) 自首する (surrender)	5651	が (ga) 逮捕する (arrest)	82112	を (wo) 逮捕する (arrest)
136	-者 person	702	署員 PS staff	15285	人 person
117	犯人 criminal	698	警察 police	8484	-者 person
96	彼 he	376	警察官 police officer	5563	男 man
68	人 person	368	-署 police station	2804	-名 name
63	男 man	230	県警 prefectural police	1188	犯人 criminal
36	-犯 crime	177	-員 -staff	1185	男性 male
30	少年 boy	153	府警 prefectural police	763	-ら -s
26	高校生 high-scool student	137	当局 authorities	671	おまえ you

Proposal2:
Selection-and-Classification Model Considering Argument Types



Conclusion

Proposal1:

A similarity feature between argument positions using distribution similarity

Proposal2:

Selection-and-Classification Model Considering Argument Types