Language and Information(2)

Jyunichi Tsujii

・The University of Tokyo Interfaculty Initiative in Information Studies Graduate School of Interdisciplinary Information Studies

- Shouldn't structured knowledge be available for all people, the same as the technologies understandable enough for everybody? MIMA search and Natural Language processing are both difficult matters

User-interface, visualization, knowledge discovery from a large quantity of data, Analysis tools and Question answering system

Understanding existing knowledge is always difficult.
It is able to promote the understand of knowledge. If you desire more, It is Intellectual laziness.

- It can be assumed that printing technology has more impact than Information technology. Can it cause a big change? Whereas the printing technology made databases possible, Information technology seems to just have accelerated the process.

One way by printing, asynchronous communication, two-way communication, synchronous・asynchronous communication

Not only accumulation and distribution, but also managing and processing.
organizing information can be possible by text as a material

- The way information came to be stored was as follows. Oral tradition → Handwriting → Printing and Publication of books → Conversion into electronic formats. One can only wonder what will come as the next stage.

・In the sense that long-term storage is not ideal. (Deterioration, Capacity)
・Not image, the importance of reading and processing information as possible remains.

- Scientists push on toward Information Technology Innovation. This caused greater risk of being used for power.

Technology is neutral.(It may be too simplicity.) Is it certain that Centralized information to a person in authority?
This enables the demagogue, but it also causes the coming change.
Interaction between real world and factitious or arbitrariness of the interpretation in information world is more important.
I think, "what is real" problem will remain.

- I think that Information technology has evolved to symptomatic treatment. Because of this overly-large amount of information, It have become useless after all.

Has information technology developed like a symptomatic treatment? Is it the cause of information overload?
We need a technology which can select valid information from too much information, or helping technology.(judgment, value...) We need a technology which decide a mutual relationship in information overload.

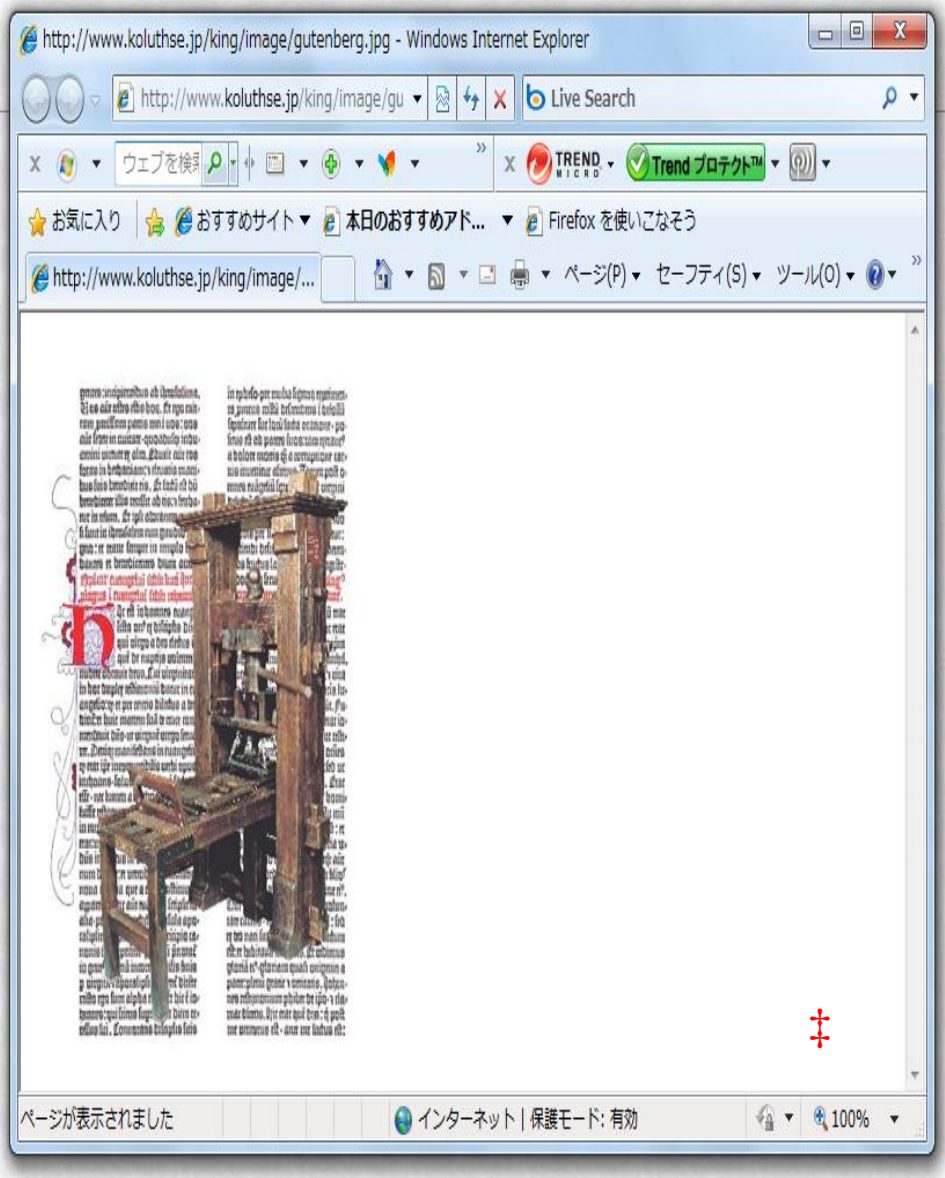
Language and（Meaning･Context･Memory･Structure･Interpretation）

- An Infant, A picture book
- Bilingual
- Justice＝to be fair
- Freedom Fighters , Terrorists
- Playing, A Structure of *Amae*

15世紀になると、聖書はグーテンベルグの発明した印刷機によって大量に、しかも正確に印刷され、さらに世界各国の言葉で印刷されるようになりました。グーテンベルグの印刷機で大量に印刷されたドイツ語やフランス語の聖書は、それまでそれぞれの地域でつかわれていた「方言」を標準化し、現代のドイツ語、フランス語の基礎ができました。こうしてグーテンベルグの印刷機が、ルネサンスなど文化の基盤を作り、近代への扉を開くことになったのです。

展示品

グーテンベルグ印刷機 (復刻機)



Comment on Tsujii's Lecture

- In the cyber world, a unique Japanese form is being formed. By it, does a situation of divided identity happen? Does Standardization of Language take language's individual style and personality?

There are some people who change their personality in the cyber world. In this case, It is possible that Language in a cyber community does not change personality. Is the relation of cause and effect opposite? Sometimes, that language changing personality is true. This may cause one to establish another character.

A development of the Internet or electronic communication promotes to "Standardization of Language" Is it true?
It is the opposite case of printing technology's standardization
If anything, it seems that diversification of language and fragmentation of Communication is occurring...

- Interpretation of Language depends on each person's background.

When we make computer do language processing, what kind of groundwork does the computer need?

At the first time of artificial intelligence: Turing test's episode. Turing's prophecy and the award of Rebner

A research on Language understanding:
Language and it's interpretation, Knowledge Representation(70's-80's)

By circumventing this problem, there is a development of technology.

- As unifying Language by a development of printing technology,

Is the way of interpretation unified by a development of language processing technology?

By Standardizing connection with Language and it's interpretation,
there is a trial to promote making efficient of communication.
A trial of Semantic Web : If you have an interest, please search out this information.

- If language is closely connected with human culture (meaning, information, knowledge), how does language background of culture deal with gathering and accumulating information?

This difficult problem will be our future challenge.
The language background is reflected in its data form, and the model needed for such data processing is gradually becoming a research target. Furthermore, we register inclinations to focus also on the technology of processing emotional expressions and evaluations based on this data, so the future is...

- Amongst various multiple languages there does not seem to exist any 1:1 semantic congruence and this offers a question how precisely are the words semantically connected in the discourse of language processing technology?

I will touch upon this issue during my lecture.

- Is there a possibility that the common daily life language could be corresponding with the Language processing technology?

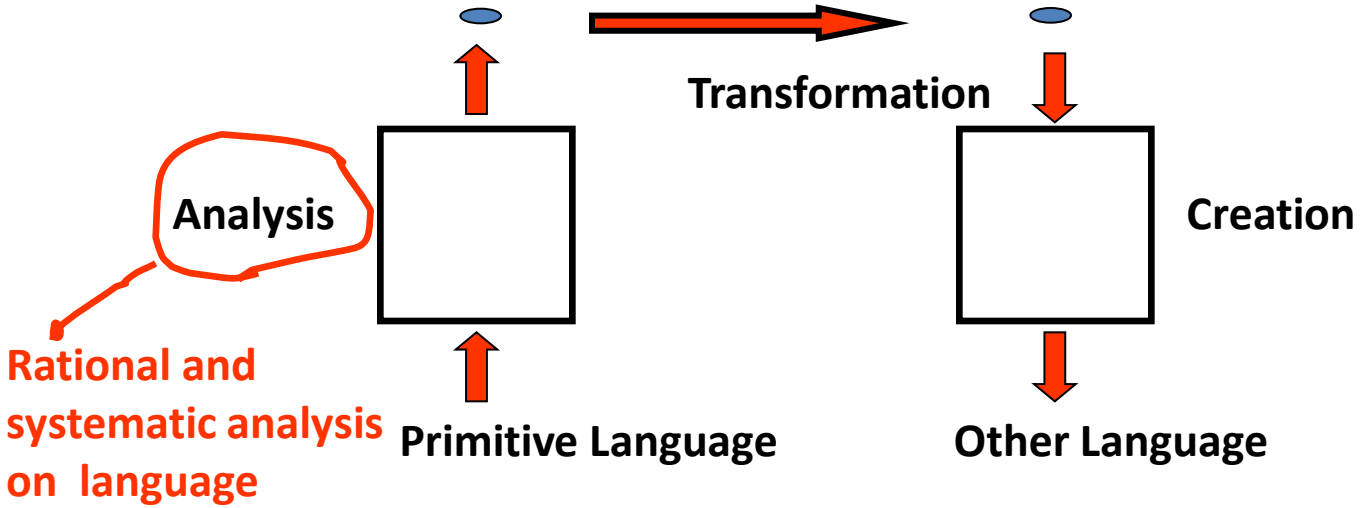
- Wouldn't it be better to have just one global unified language?

Reasons for this vision being troublesome are the diversity of thoughts expressed by words, the dynamic nature of language, the problem of newly-coined words, A problem of technical terminology, the obstacle with words written in Katakana.

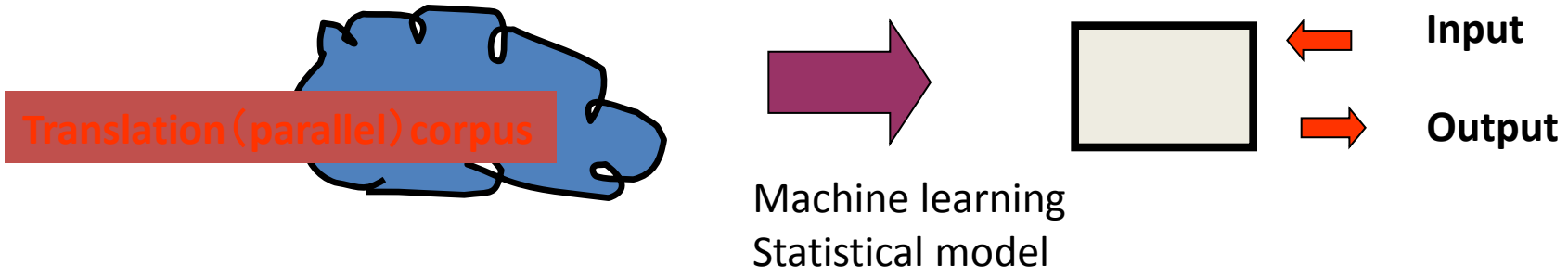
TECHNOLOGY OF LANGUAGE PROCESSING :
A example of machine translation

An interest for machine translation of "Minority Language"

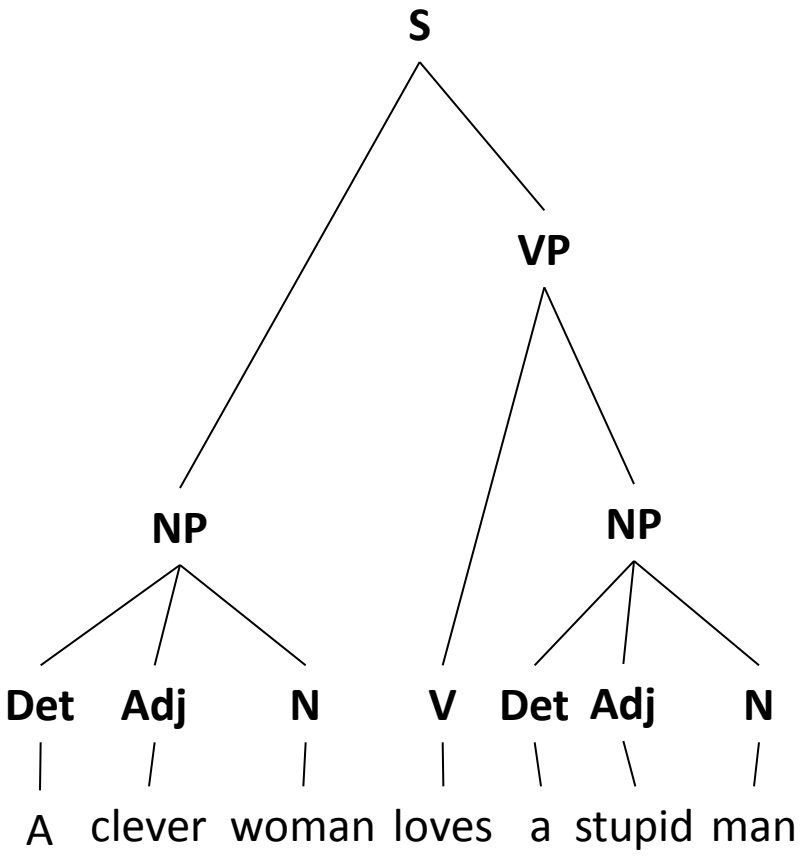
（1） Machine translation of rationalism： Translation which is based on a rule

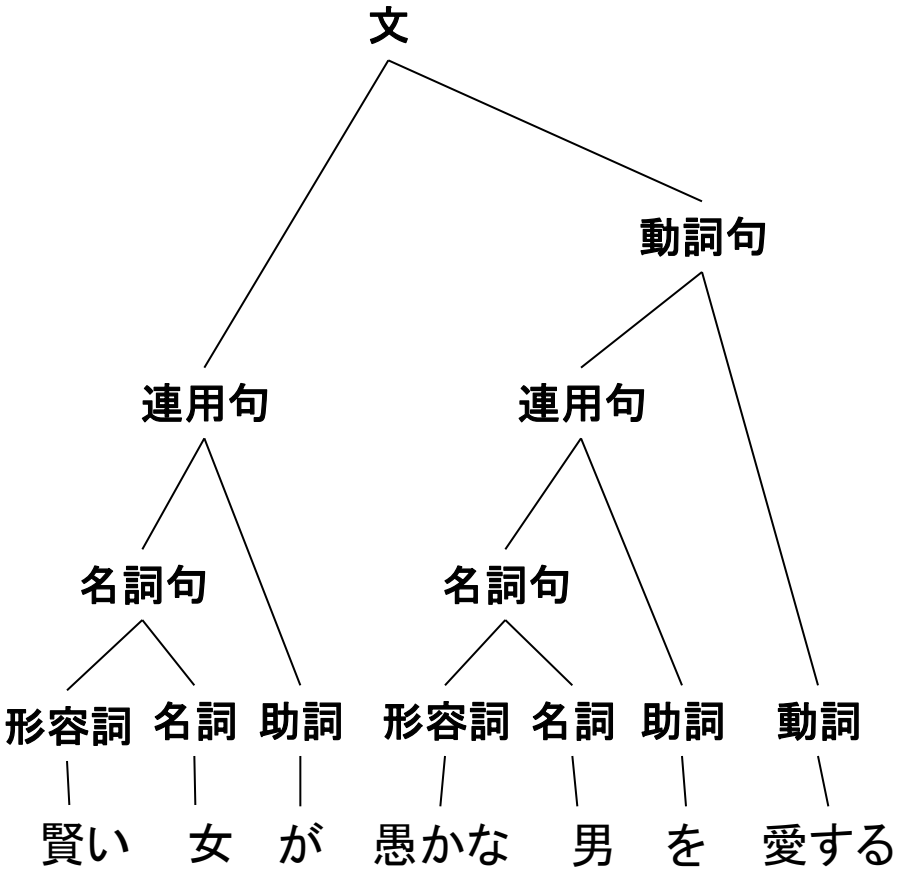


（2） Empiricism machine translation： Translation which is based on a rule

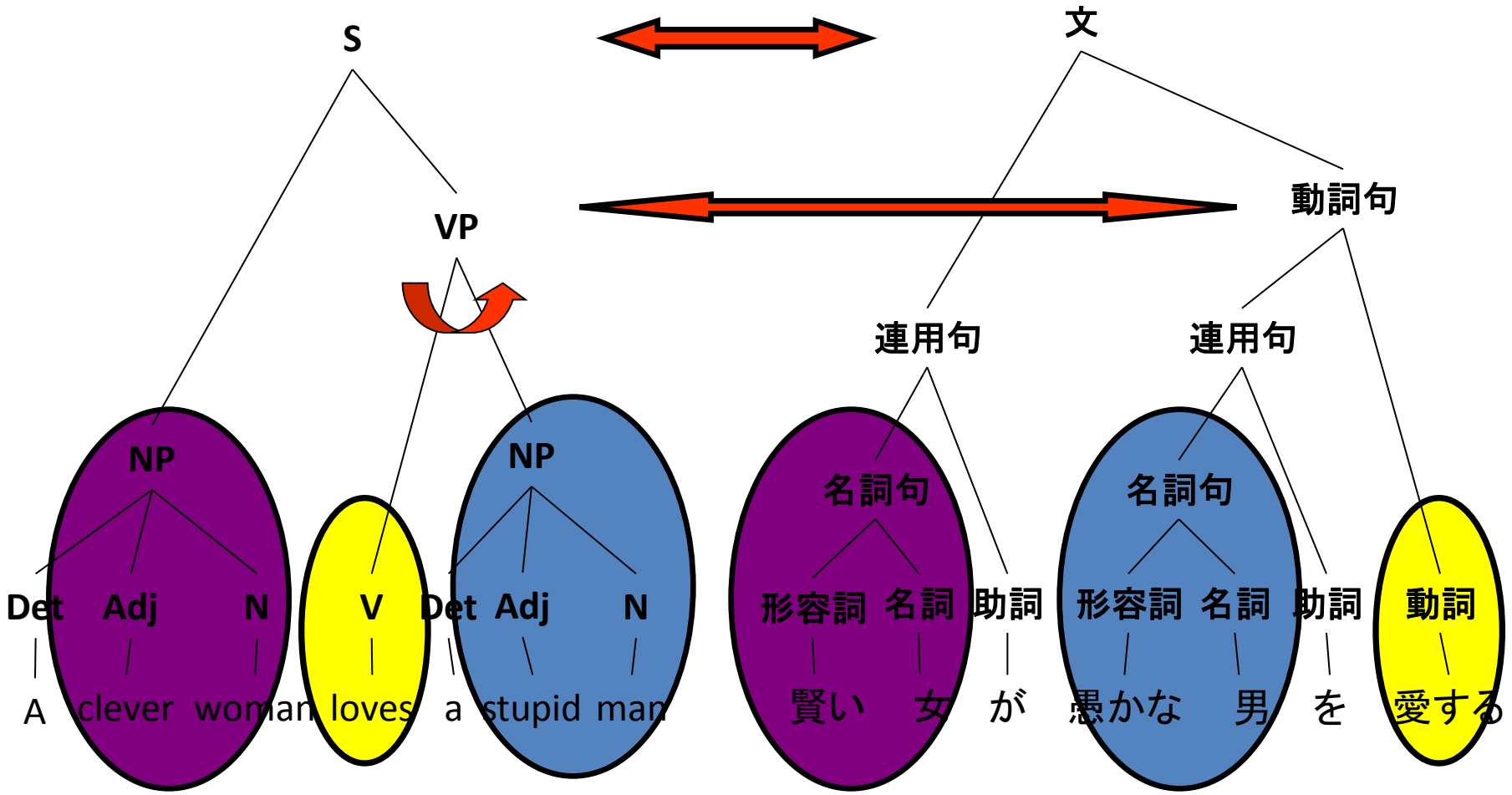


Constitutive Translation



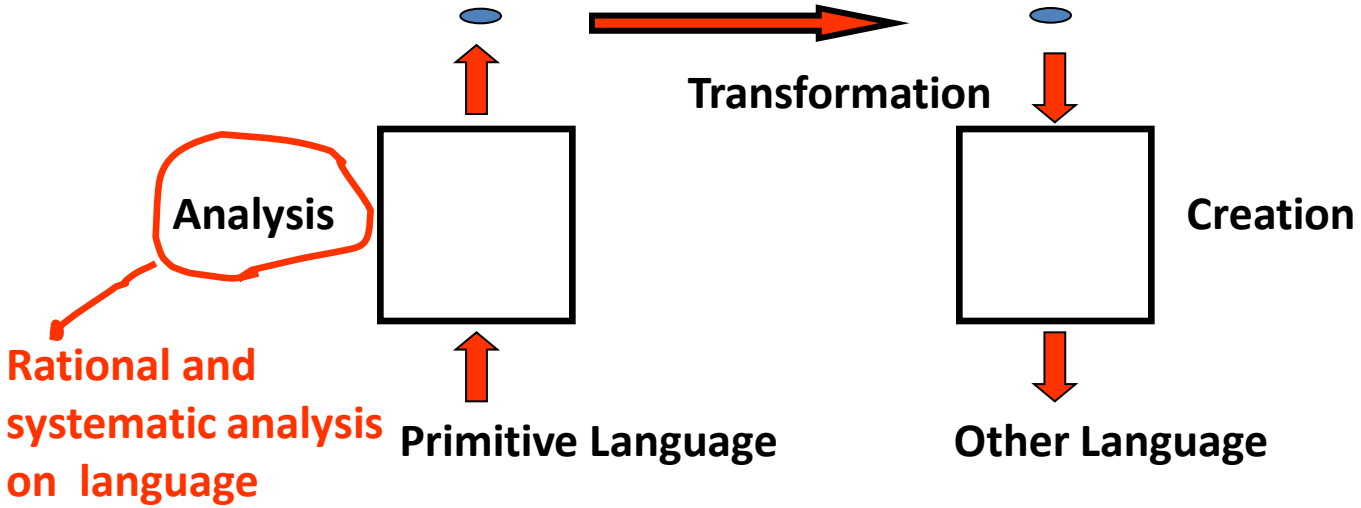


Constitutive Translation

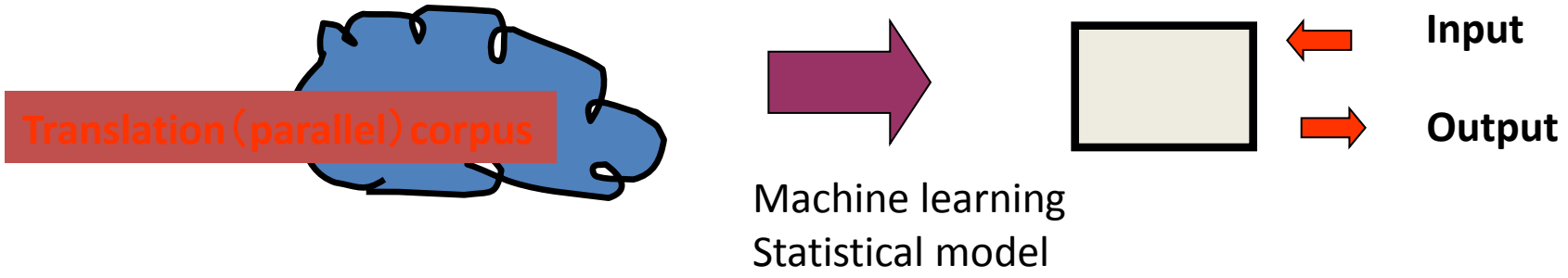


An interest for machine translation of "Minority Language"

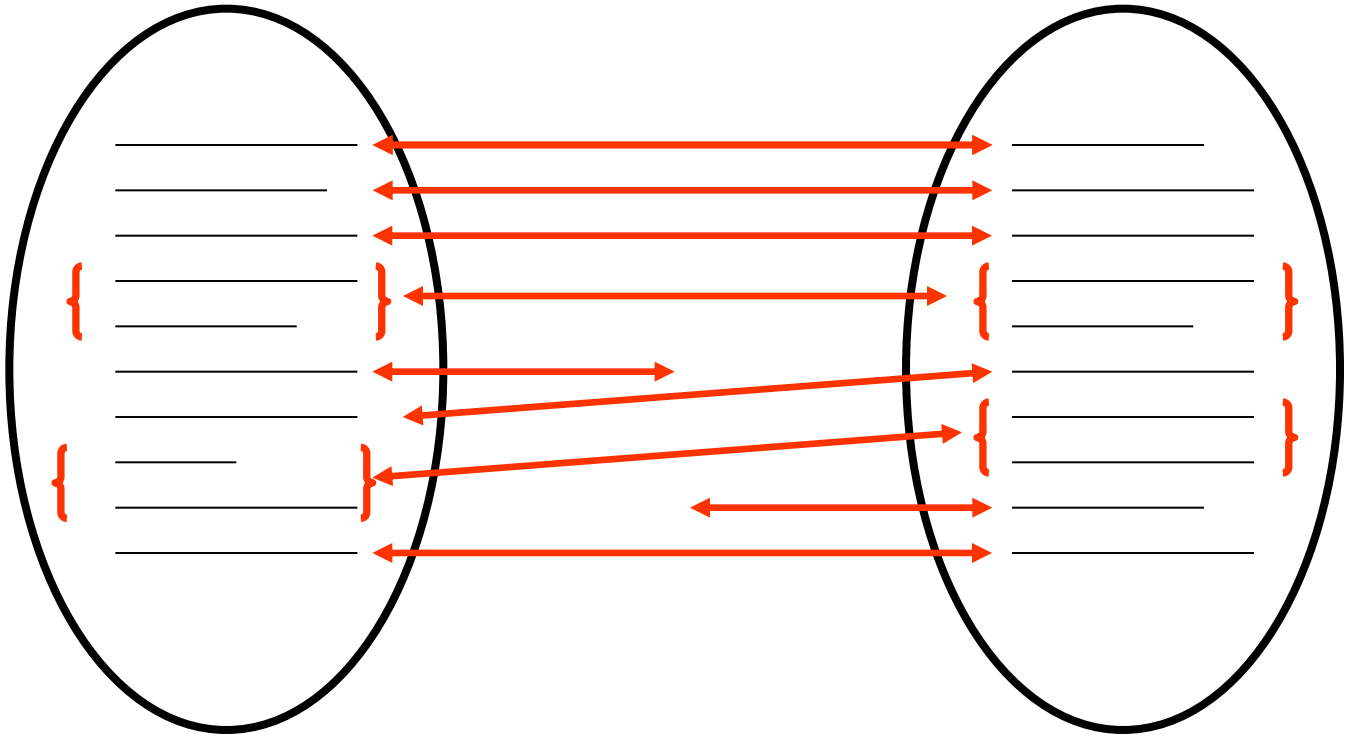
（1） Machine translation of rationalism： Translation which is based on a rule



（2） Empiricism machine translation： Translation which is based on a rule



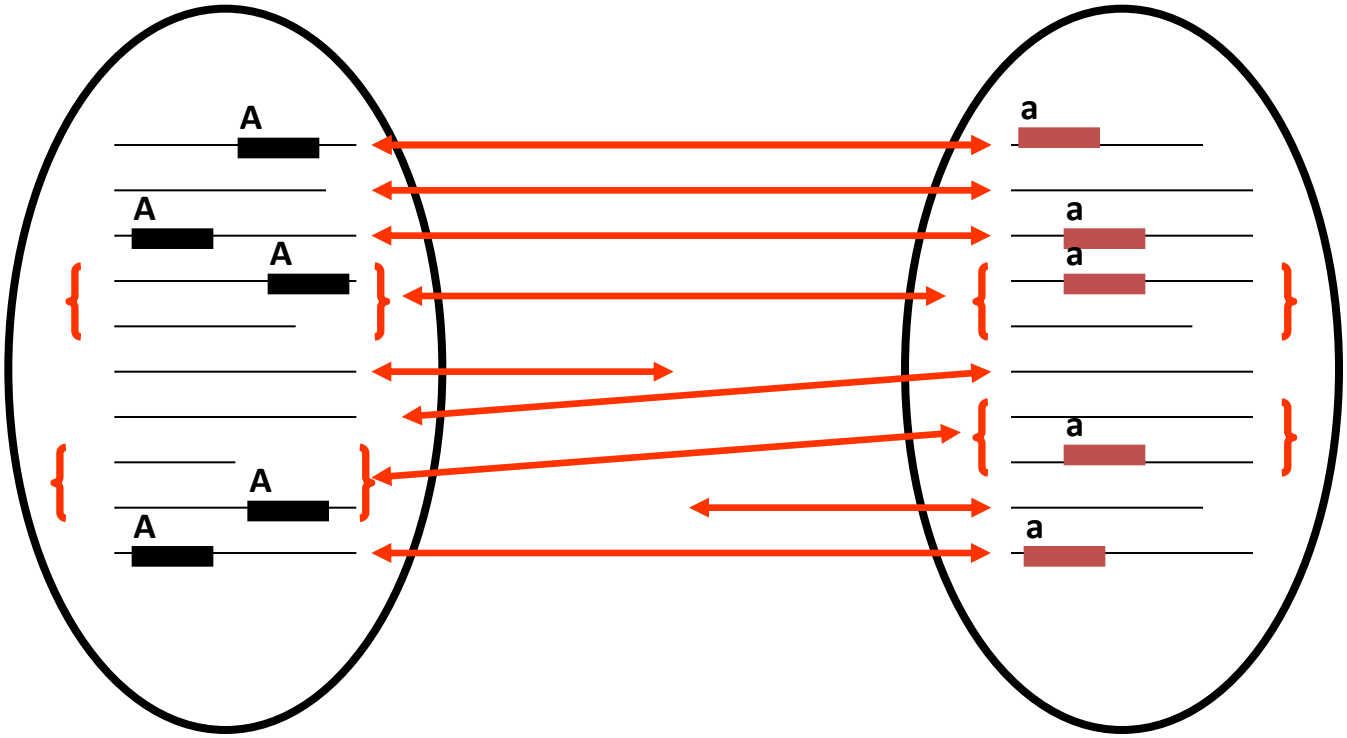
Alignment of Sentence



Japanese English

Alignment of words

Alignment of sentence

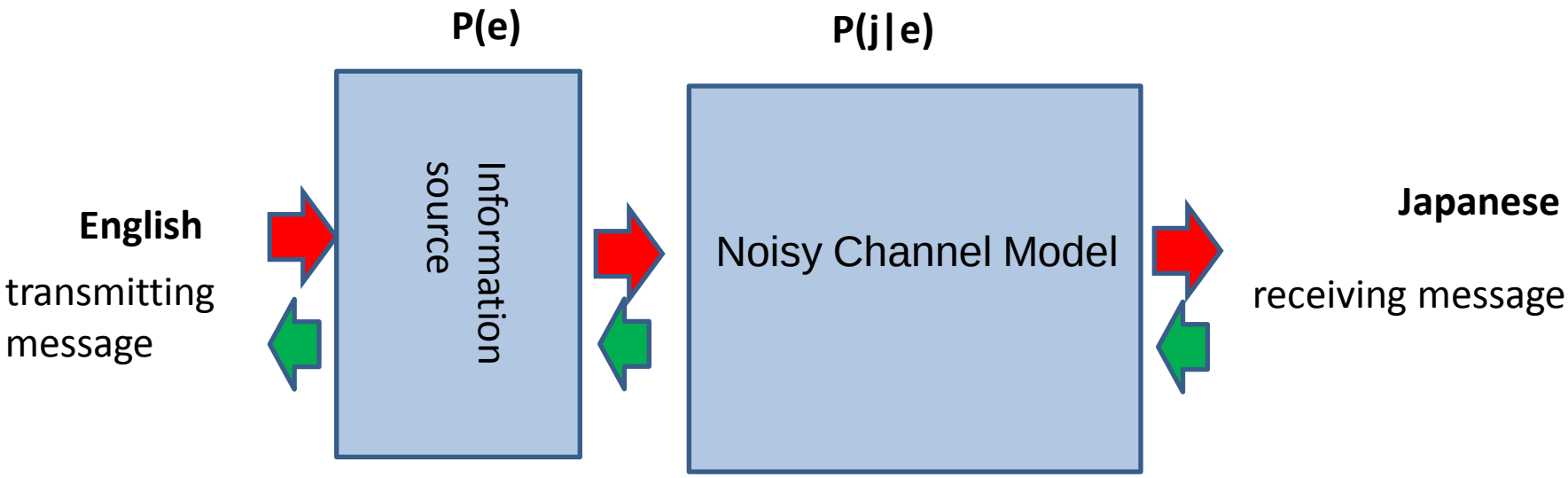


Japanese English

Alignment of words

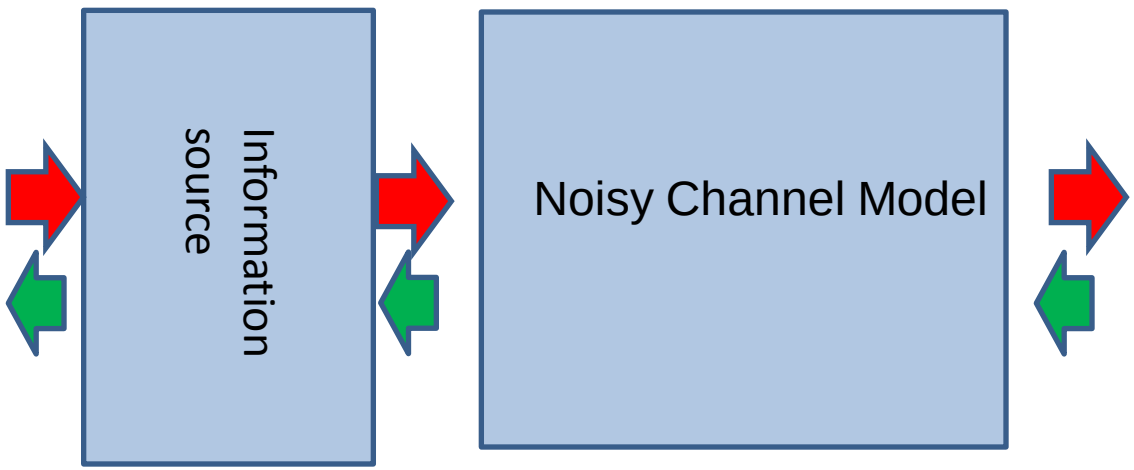
A Statistical Model of Translation : $\underset{e \in E}{\text{ARGMAX}} \{ P(e)P(j|e) \}$

Noisy Channel Model



A Statistical Model of Translation : ARGMAX { P(e)P(j|e) }

Noisy Channel Model



e

transmitting message

j

receiving message

Spring has come

春が来た

P(e)
A probability of appearing "Spring has come" in English text

≒

P(spring|*)P(has|*, spring)P(come|spring, has)

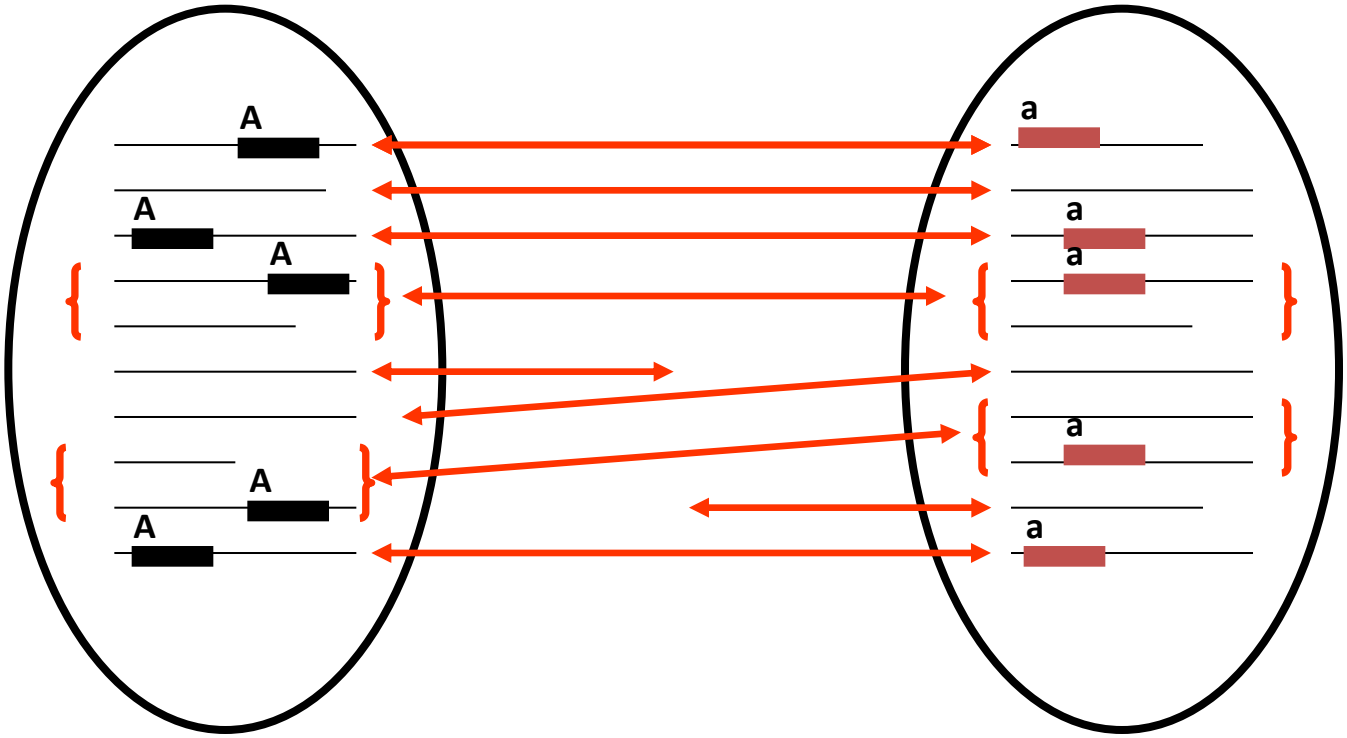
P(j|e)
A probability of "Spring has come =「春が来た」"

≒

P（春|spring)P(くる|Come)P(た|has)

A Statistical Model of Translation： ARGMAX { P(e)P(j|e) }

Alignment of sentence



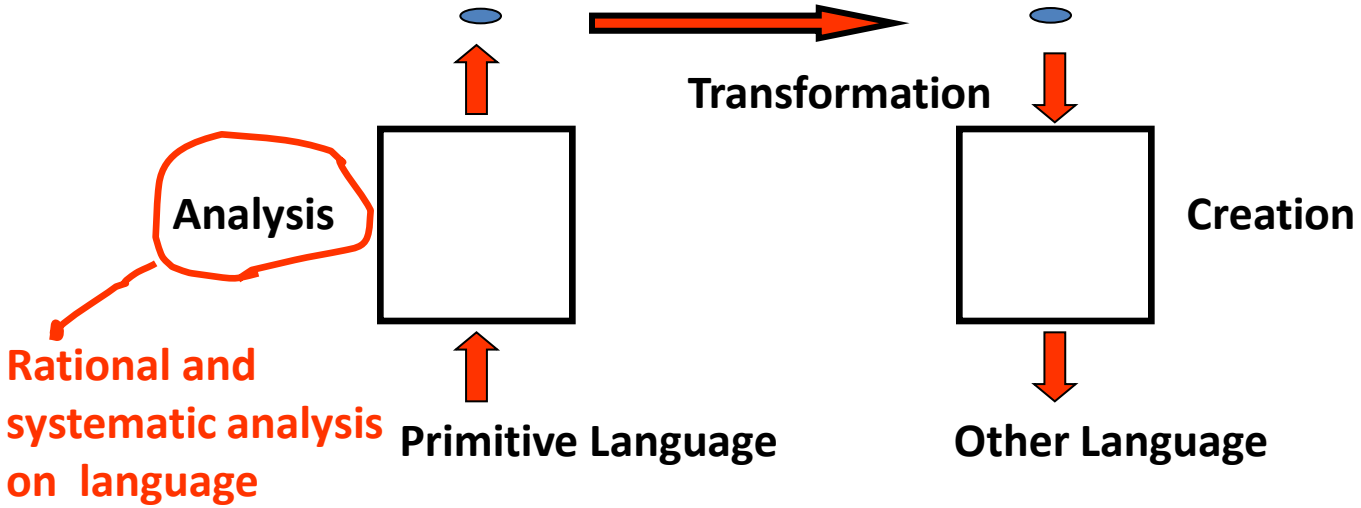
Japanese English

Alignment of words

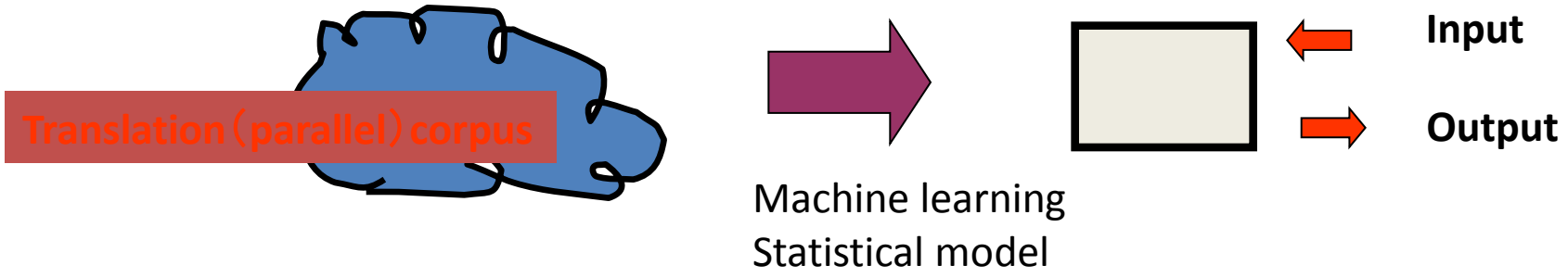
A Statistical Model of Translation : $\underset{e \in E}{\text{ARGMAX}}$ { P(e)P(j|e) }

An interest for machine translation of "Minority Language"

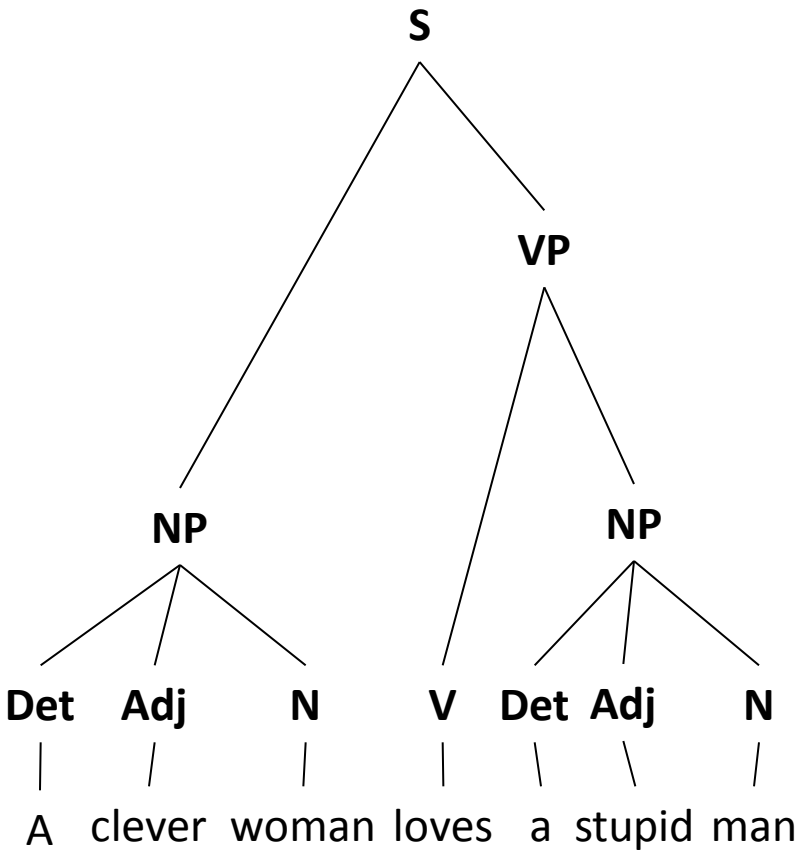
（1） Machine translation of rationalism： Translation which is based on a rule

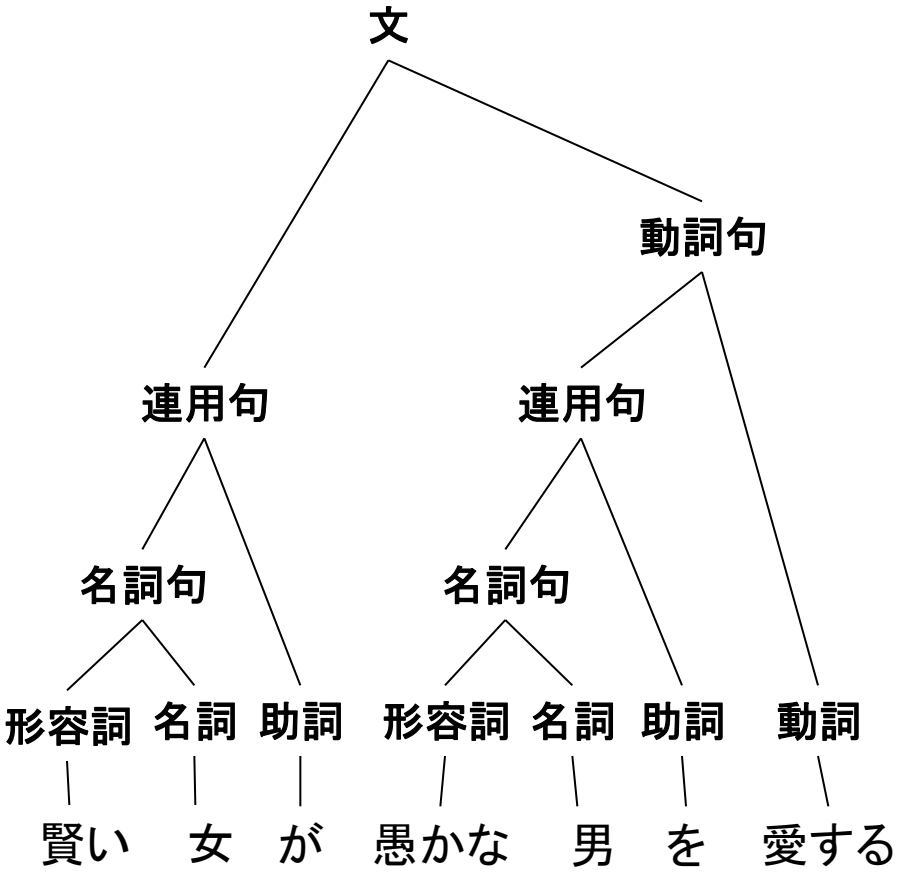


（2） Empiricism machine translation： Translation which is based on a rule



Constitutive Translation





Regularity of Language

- Syntactic Regularity — Syntax

  Regularity on Language form
  - Morphologic Regularity
  - Syntactic Regularity

- Semantic Regularity

  Regularity between Language and expression

- Pragmatic Regularity — Semantics

  Language, expression, Regularity in an Environment（User, Listener）

Infinity

Pragmatics

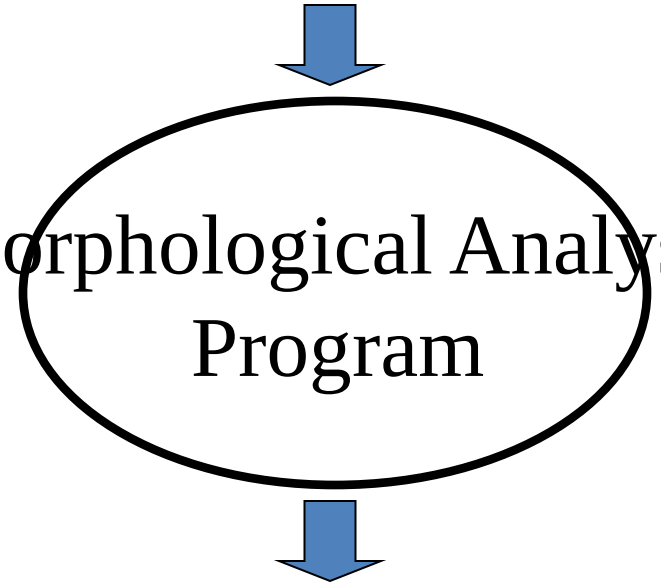
Morphological Regularity (Morphology)

- Inflection, Conjunctive Relation
- 遊ぶ—遊ば ない (未然） Asob a
  遊び ます（連用） Asob i
  遊ぶ とき（連体） Asob u
  遊ぶ (終止） Asob u
  遊べ ば （仮定） Asob e
  遊べ (命令） Asob e

  Stem the ending of a word

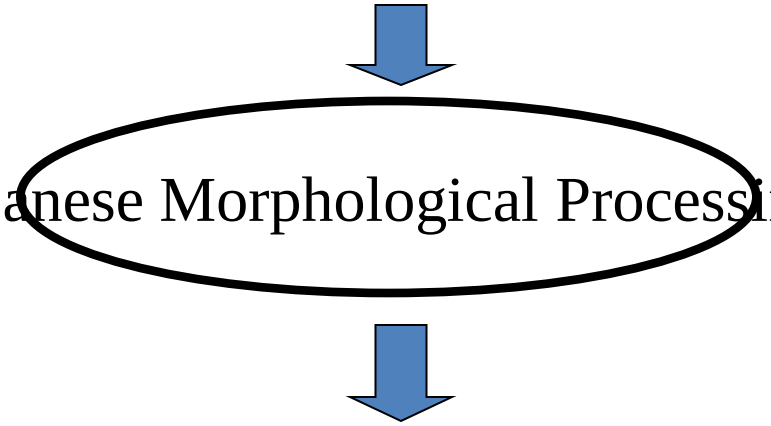
Morphological Processing of Japanese

私は計算言語学の国際学会に
代理出席しなければなりません。



私･は･計算･言語･学･の･国際･学会･に･
代理･出席し･なけれ･ば･なり･ませ･ん。

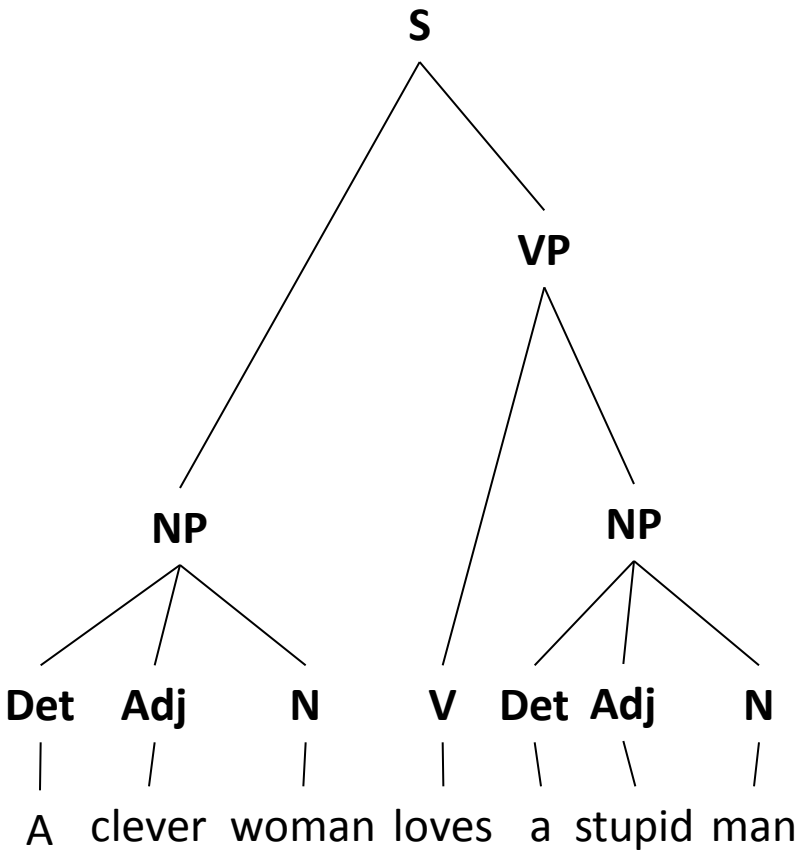
ここではきものを脱いでください。

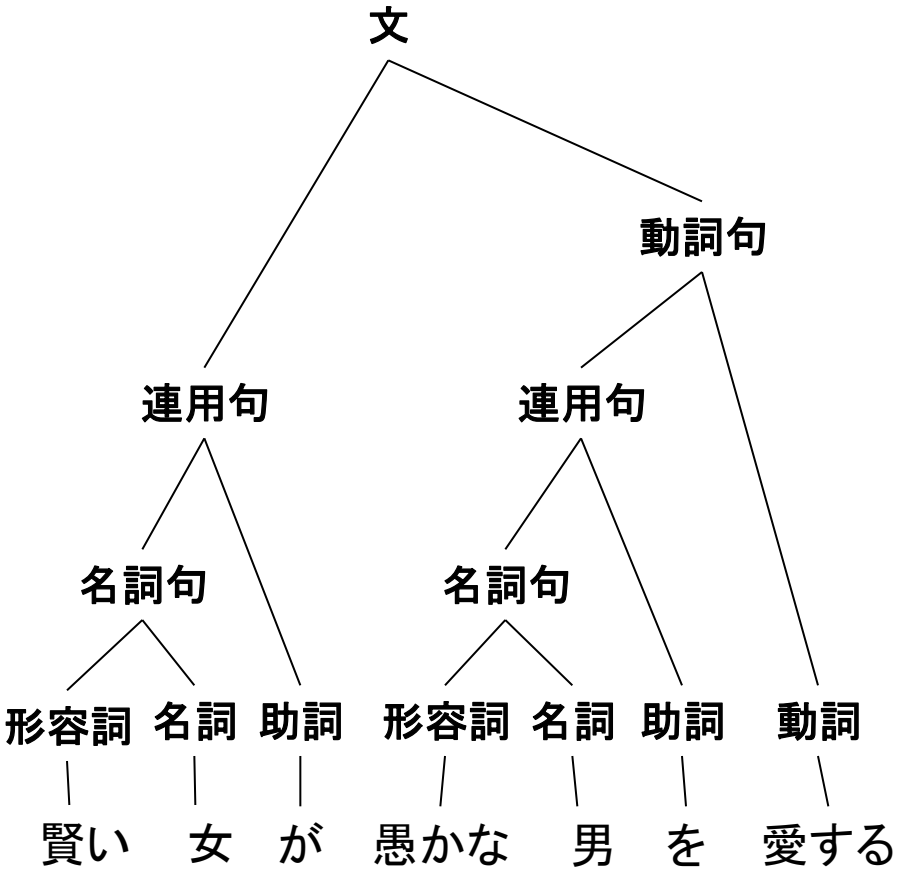


ここ･で･は･きもの･を。。。。。。

ここ･で･はきもの･を。。。。。。

Constitutive Translation

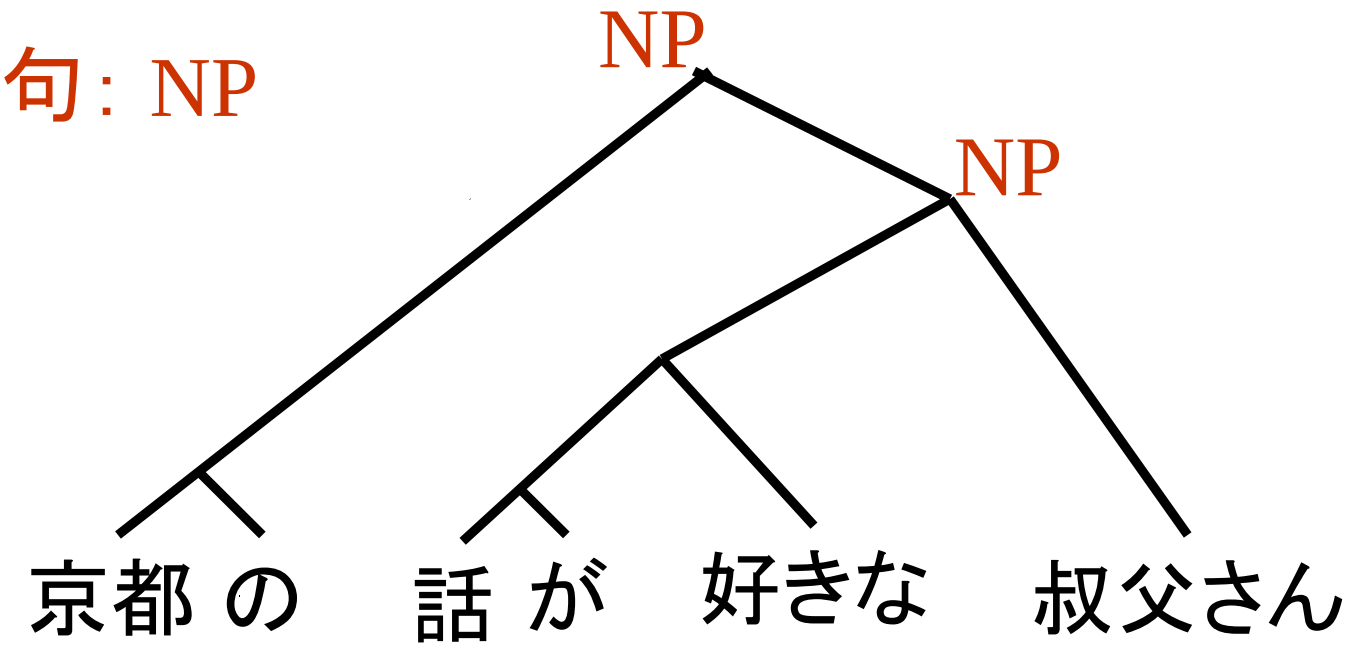


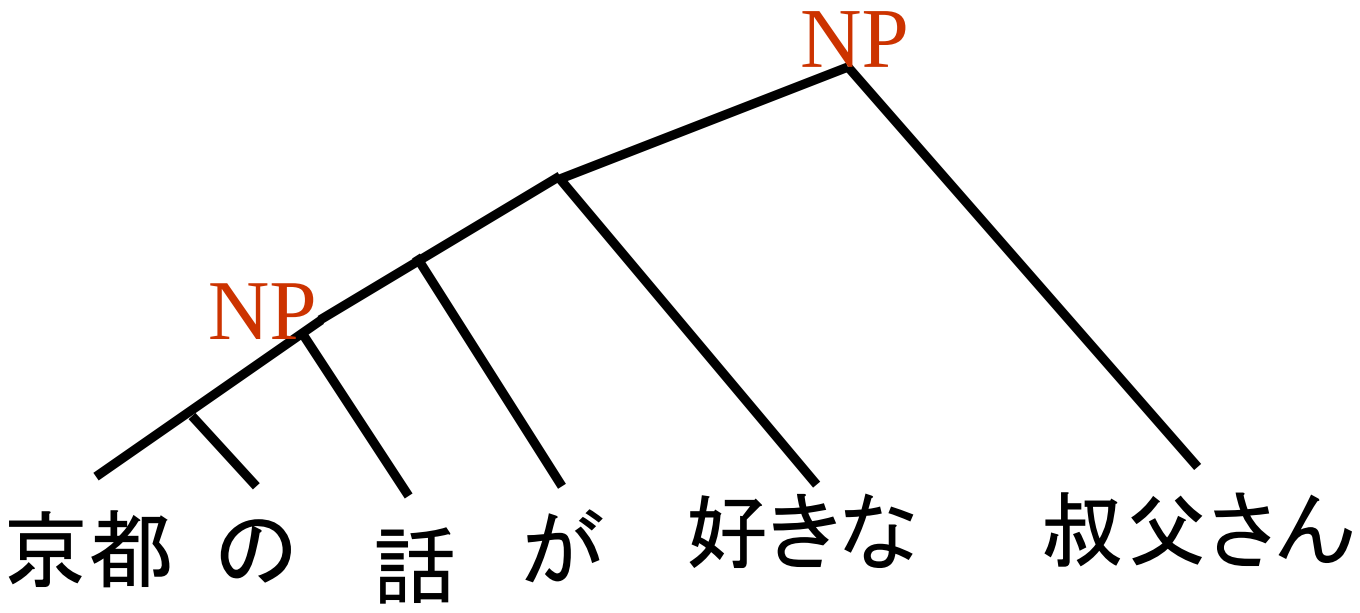


The Structure of Language

- 京都の話が好きな叔父さん
- （京都の（（話が 好きな）叔父さん））
- （（（京都の 話）が 好きな）叔父さん）

名詞句: NP

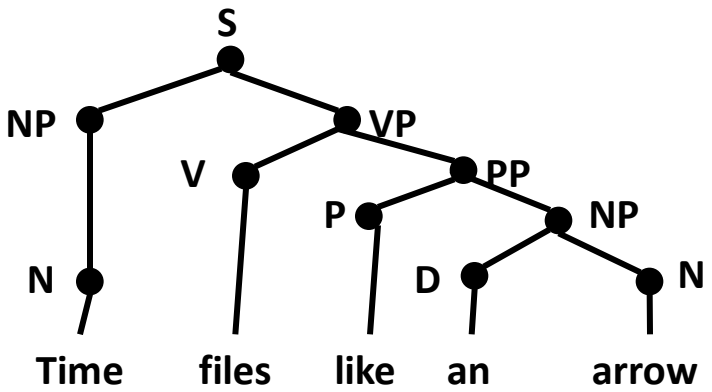




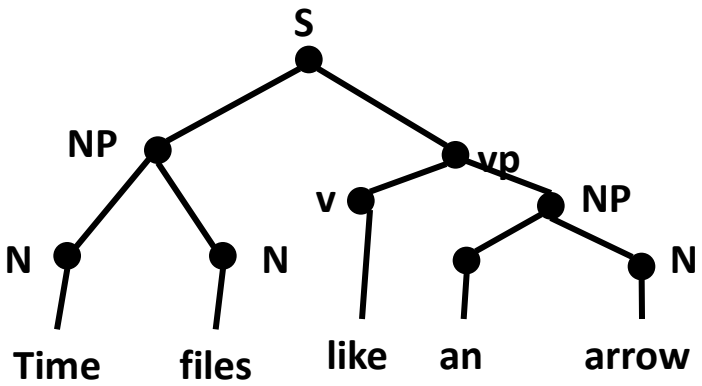
Morphological Analysis（POS Tagger)
by HMM（Hidden　Morkov　Model）

Time flies like an arrow.

N　　N　　V　　D　　N
V　　V　　P

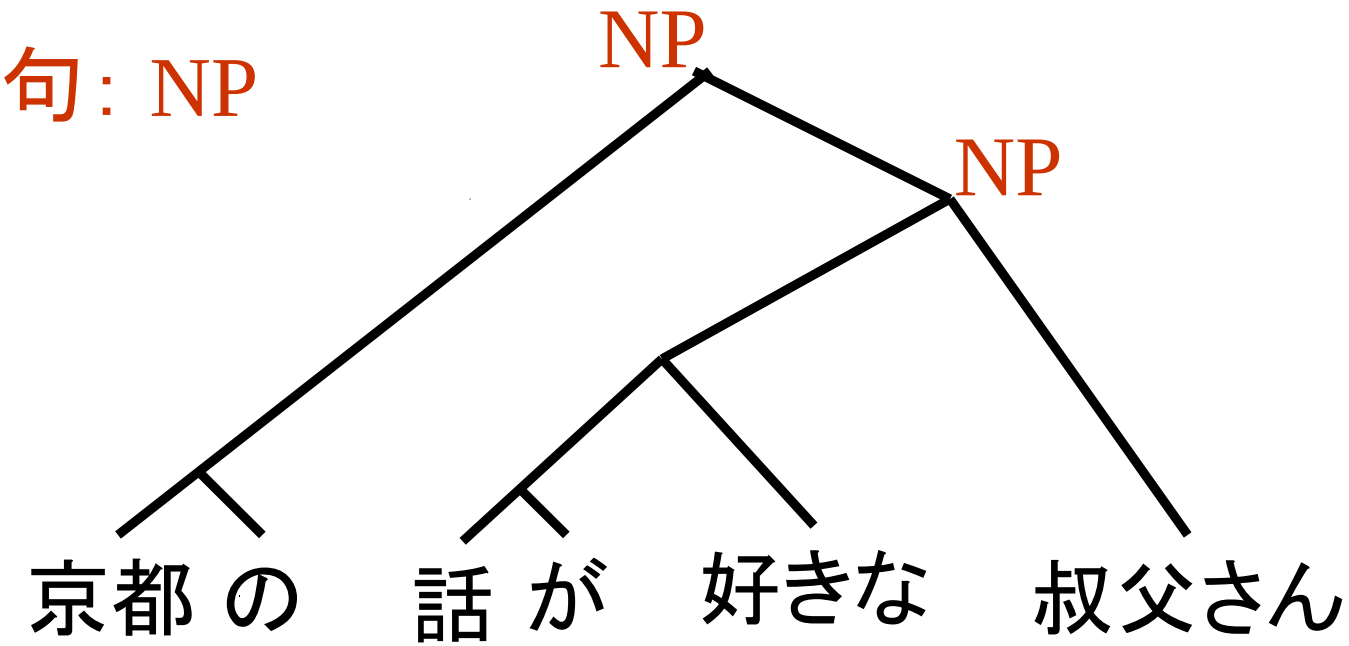


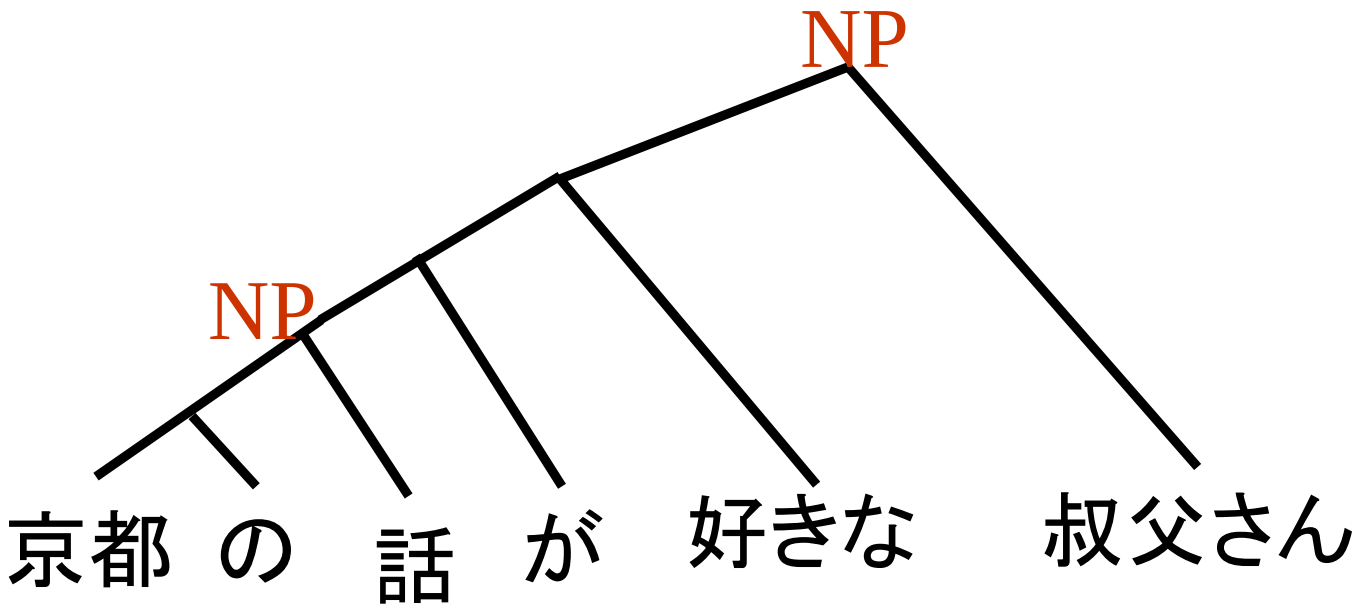
N　　V　　P　D　　N



N　　N　　V　D　　N

名詞句: NP





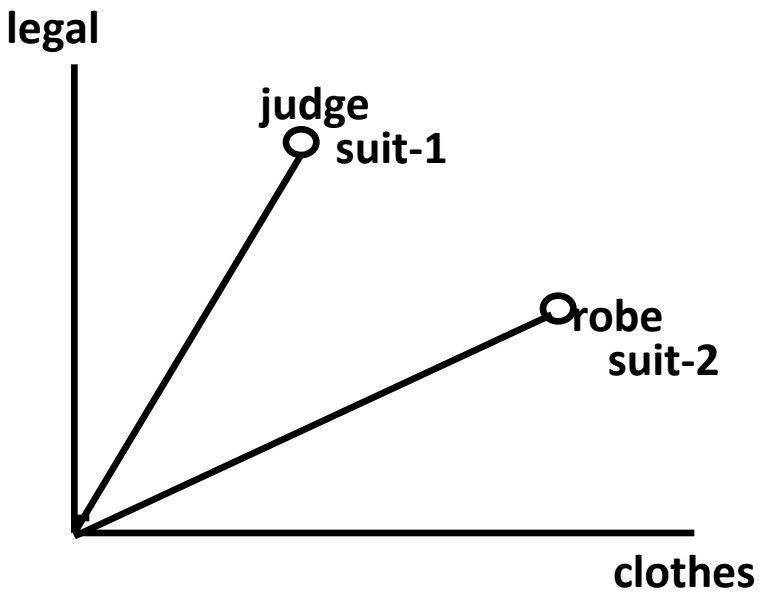
Lexical Ambiguity and Statistical Model

- Spring: 春、バネ、泉、。。。
- Bank: 銀行、川岸、。。。。

Semantic space of words

Meaning of a word is defined by Co-occuring words.

単語 / 次元	suit-1	suit-2
	judge	robe
legal	300	133
clothes	75	200



Language and（Meaning･Context･Memory･Structure･Interpretation）

- An Infant, A picture book
- Bilingual
- Justice＝to be fair
- Freedom Fighters , Terrorists
- Playing, A Structure of *Amae*

A Spaghetti Problem

- スパゲティ、スパゲッティ, スパゲッティー, スパゲッテイ, スパゲティー, スパゲテイ,スパゲッティ, ….

Microsoft Academic Search Beta

Explore over 5,381,371 papers, 155,073 were added last week.

tsujii

Advanced Search

Papers | Authors | Conferences | Journals

Jun'ichi Tsujii

Publications: 84 | Citations: 359 | G-Index: 17 | H-Index: 10

Research Interest: Natural Language & Speech, Bioinformatics and Computational Biology, Machine Learning and Pattern Recognition

University of Manchester, Manchester, United Kingdom

Bing

Freedom Fighters and Terrorists

eck author trend. Please enable it in your browser or click here to install.

Papers | Citations

Year 2006

K. Masuda, T. Ninomiya, Y. Miyao, T. Ohta, **J. Tsujii** : Nested region algebra extended with variables , 2006 (Citations: 1)

Year 2005

Akane Yakushiji, Yusuke Miyao, Yuka Tateisi, **Jun'ichi Tsujii** : Biomedical information ex-traction with predicate-argument structure patterns , 2005 (Citations: 13)

Takashi Ninomiya, Yoshimasa Tsuruoka, Yusuke Miyao, **Jun'ichi Tsujii** : Efficacy of beam threshold-ing , 2005 (Citations: 1)

Year 2004