

Committee-based Decision Making in Probabilistic Partial Parsing

INUI Takashi * and **INUI Kentaro** * †

* Kyushu Institute of Technology

† PRESTO, Japan Science and Technology Corporation

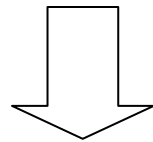
Background



- ❑ Increasing availability of large tree banks
- ❑ Success of statistical approaches to parsing

However,

Improvements appear to be getting saturated



❑ Two new directions for extending the current probabilistic parsing techniques,

- ❖ *Probabilistic Partial Parsing*
- ❖ *Committee-based decision making*

Overview of today's talk

❑ Probabilistic partial parsing

- ❖ A probabilistic extension of partial parsing

❑ Committee-based probabilistic partial parsing

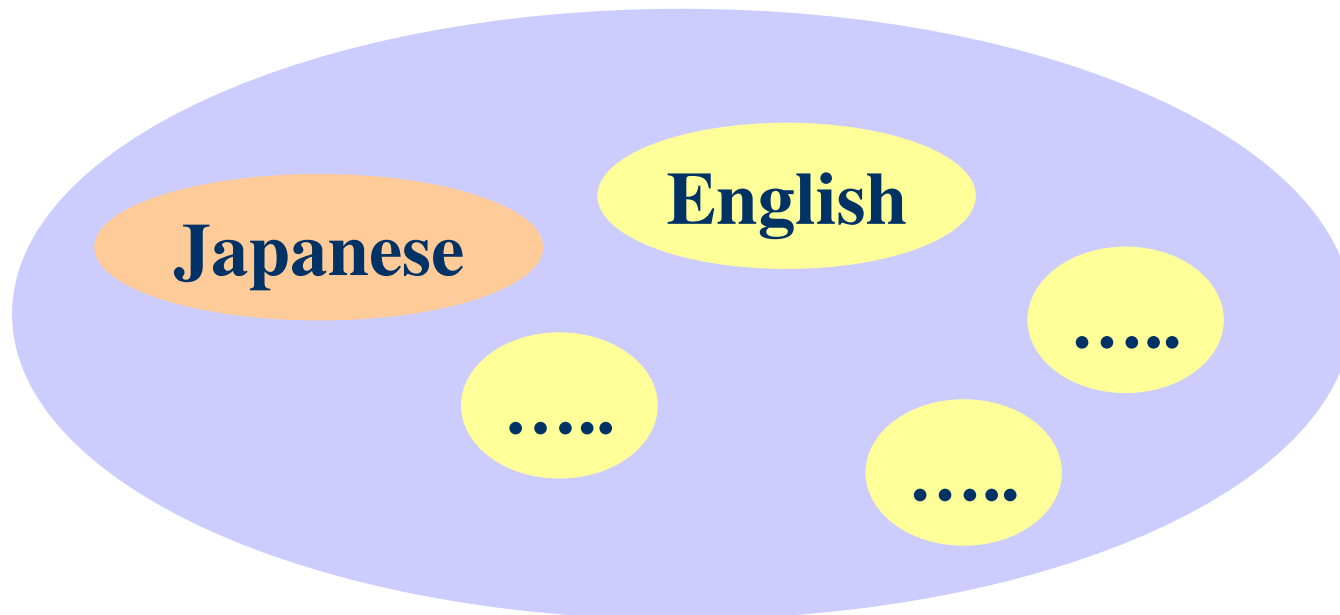
1. Probabilistic voting
2. Standardization
3. Multiple voting

❑ Experiments

Bunsetsu phrase (BP)

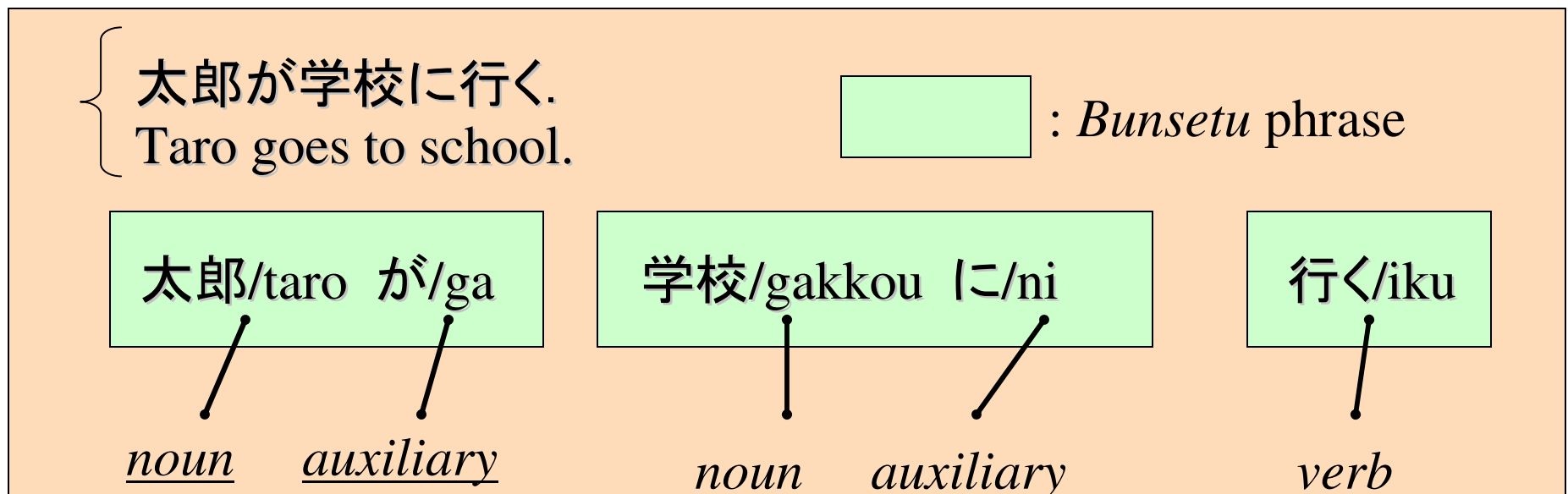
In this talk

- ❑ The target language of experiments is Japanese
- ⇒ However,
- ❑ Our proposal is **not limited to Japanese**
- ❑ It should be able to be **applied to other languages like English**



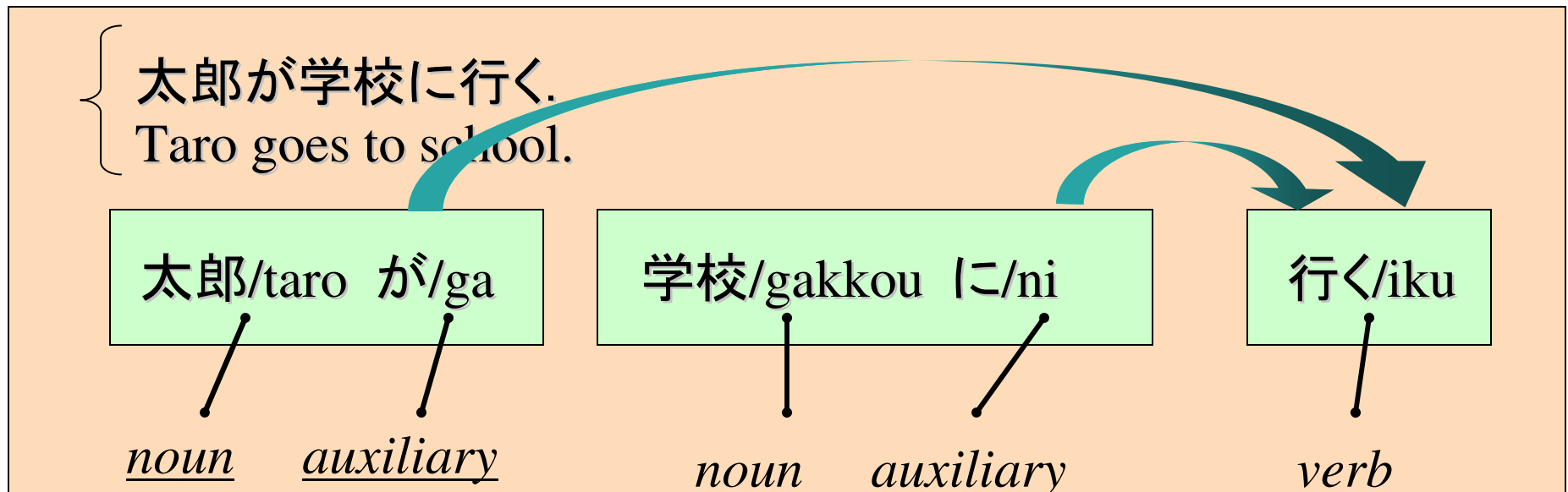
Bunsetsu phrase (BP)

- A **Bunsetsu phrase (BP)** is a chunk of words consisting of a content word (noun, verb, etc.) accompanied by some functional words (particle, auxiliary, etc.)
- A Japanese sentence can be analyzed as a sequence of BPs, which constitute an **inter-BP dependency structure**



Bunsetsu phrase (BP)

- A **Bunsetsu phrase (BP)** is a chunk of words consisting of a content word (noun, verb, etc.) accompanied by some functional words (particle, auxiliary, etc.)
- A Japanese sentence can be analyzed as a sequence of BPs, which constitute an **inter-BP dependency structure**

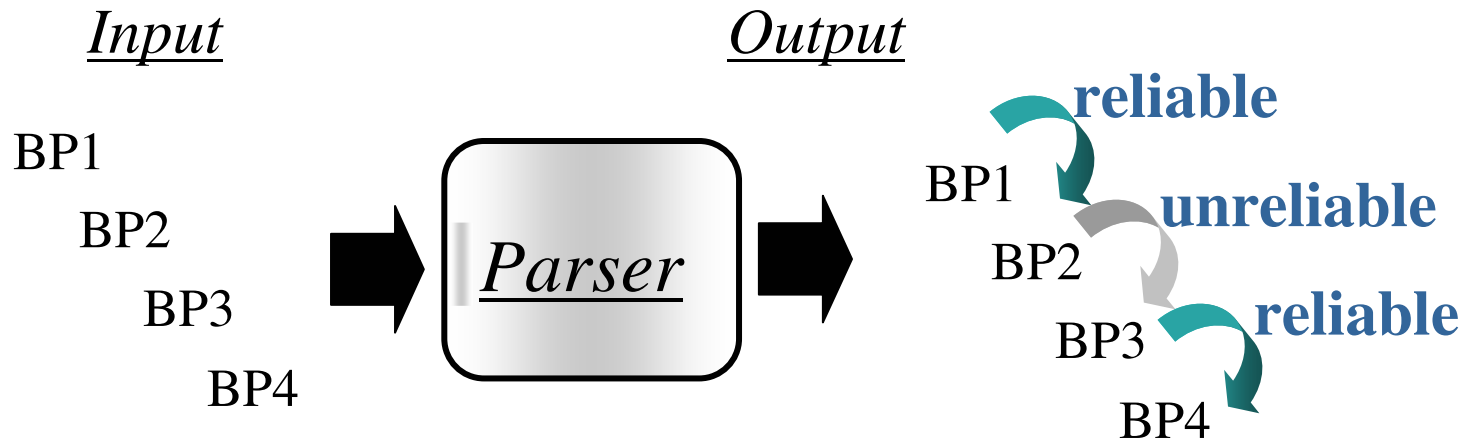


Probabilistic partial parsing

Overview

Probabilistic partial parsing

- ❑ **Probabilistic extension** (Jensen et al.,1993)
- ❑ **Output only a part of the parse tree that are probabilistically highly reliable**

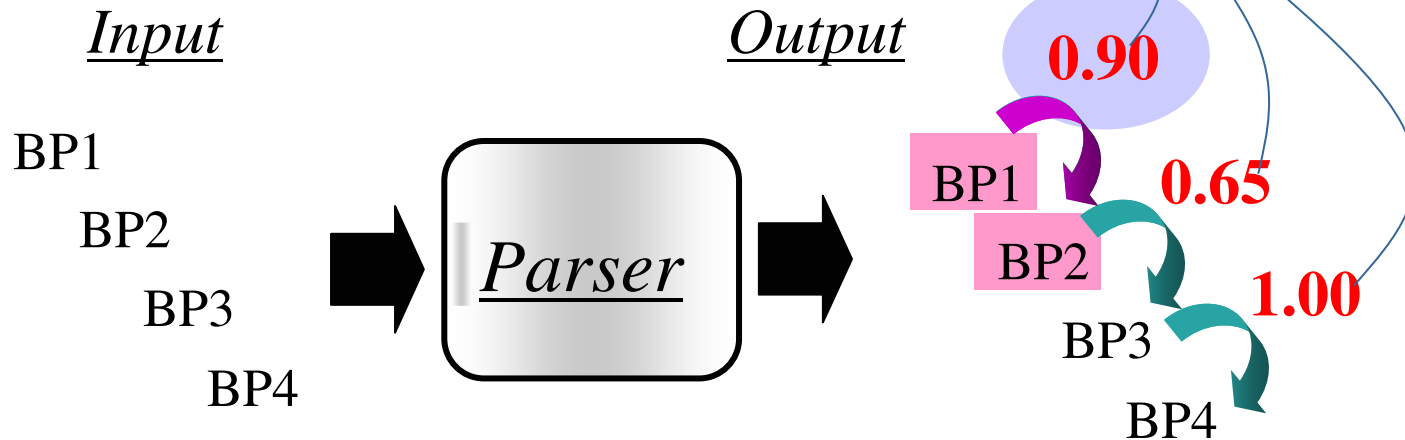


Overview

Probabilistic partial parsing

- ❑ Probabilistic extension (Jensen et al.,1993)
- ❑ Output only a part of the parse tree that are probabilistically highly reliable

Dependency probabilities (DPs)

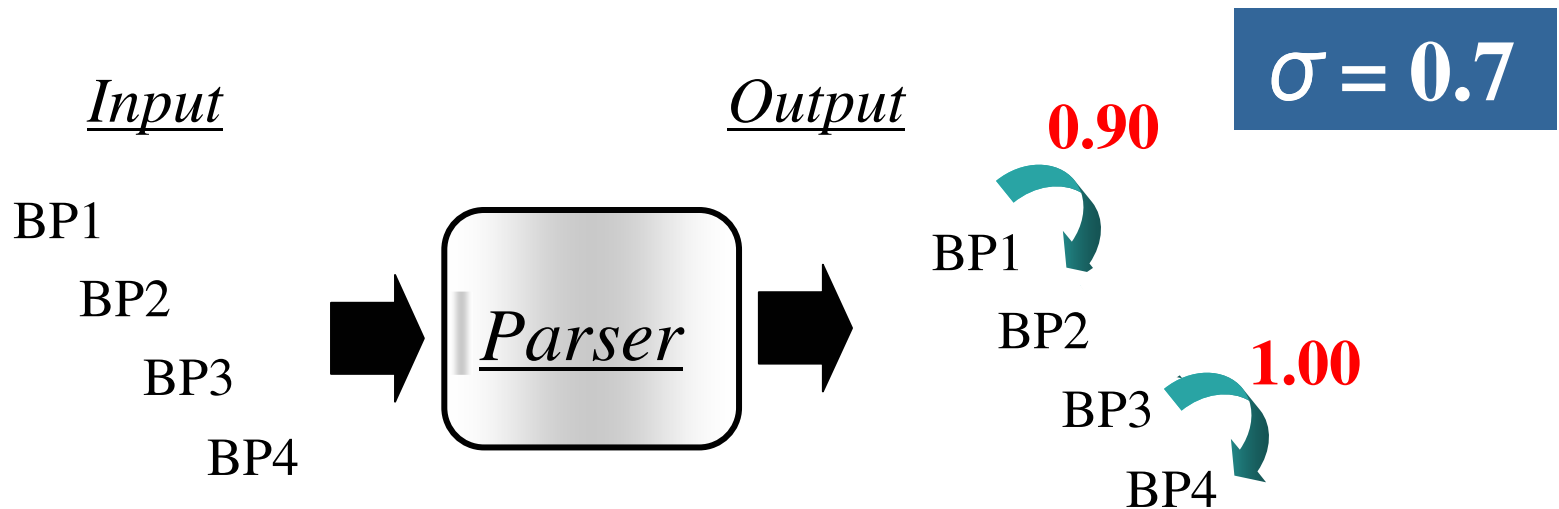


Selecting only dependency relations whose estimated probability is higher than a certain threshold σ .

Overview

Probabilistic partial parsing

- ❑ Probabilistic extension (Jensen et al.,1993)
- ❑ Output only a part of the parse tree that are probabilistically highly reliable

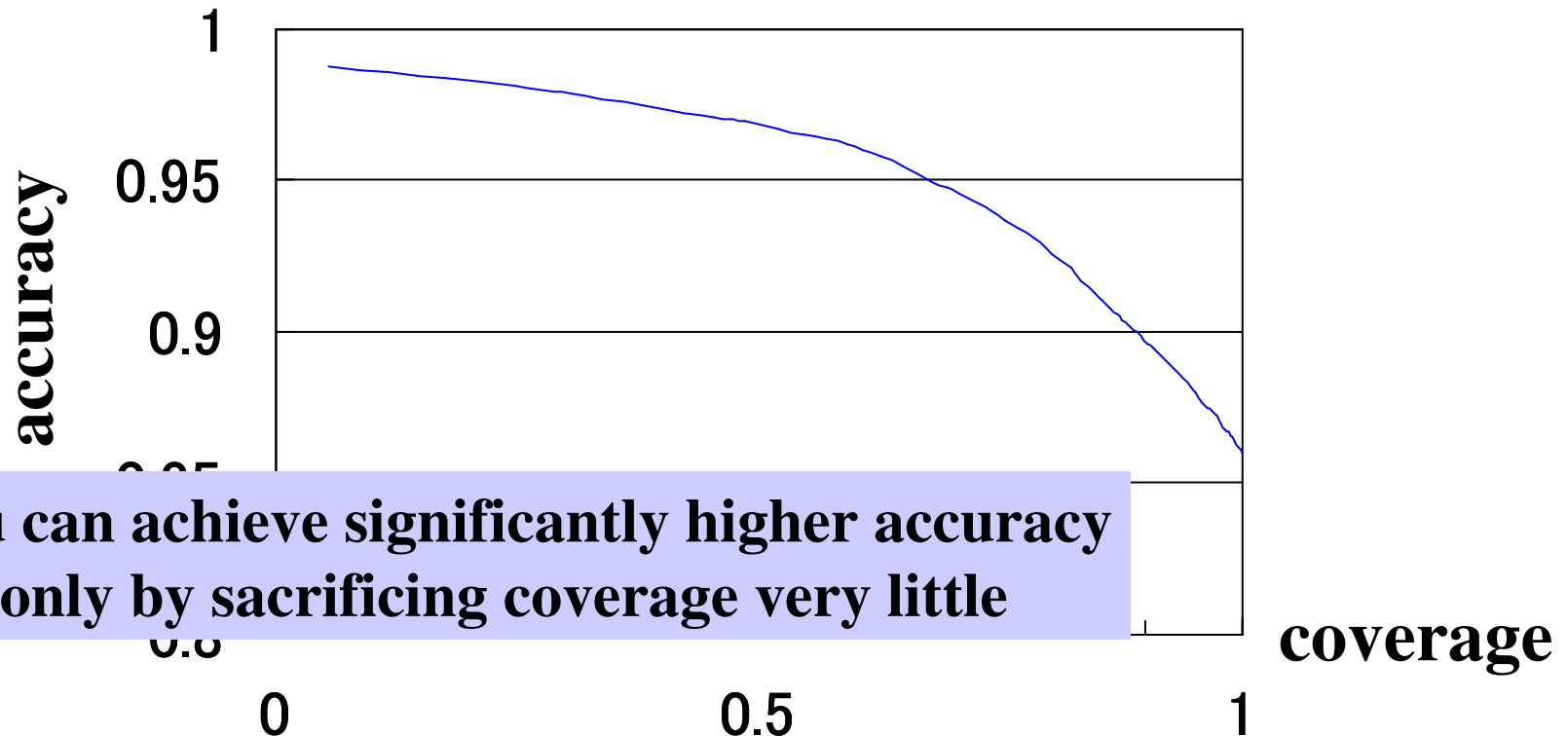


Selecting only dependency relations whose estimated probability is higher than a certain threshold σ .

C-A curve

Probabilistic partial parsing

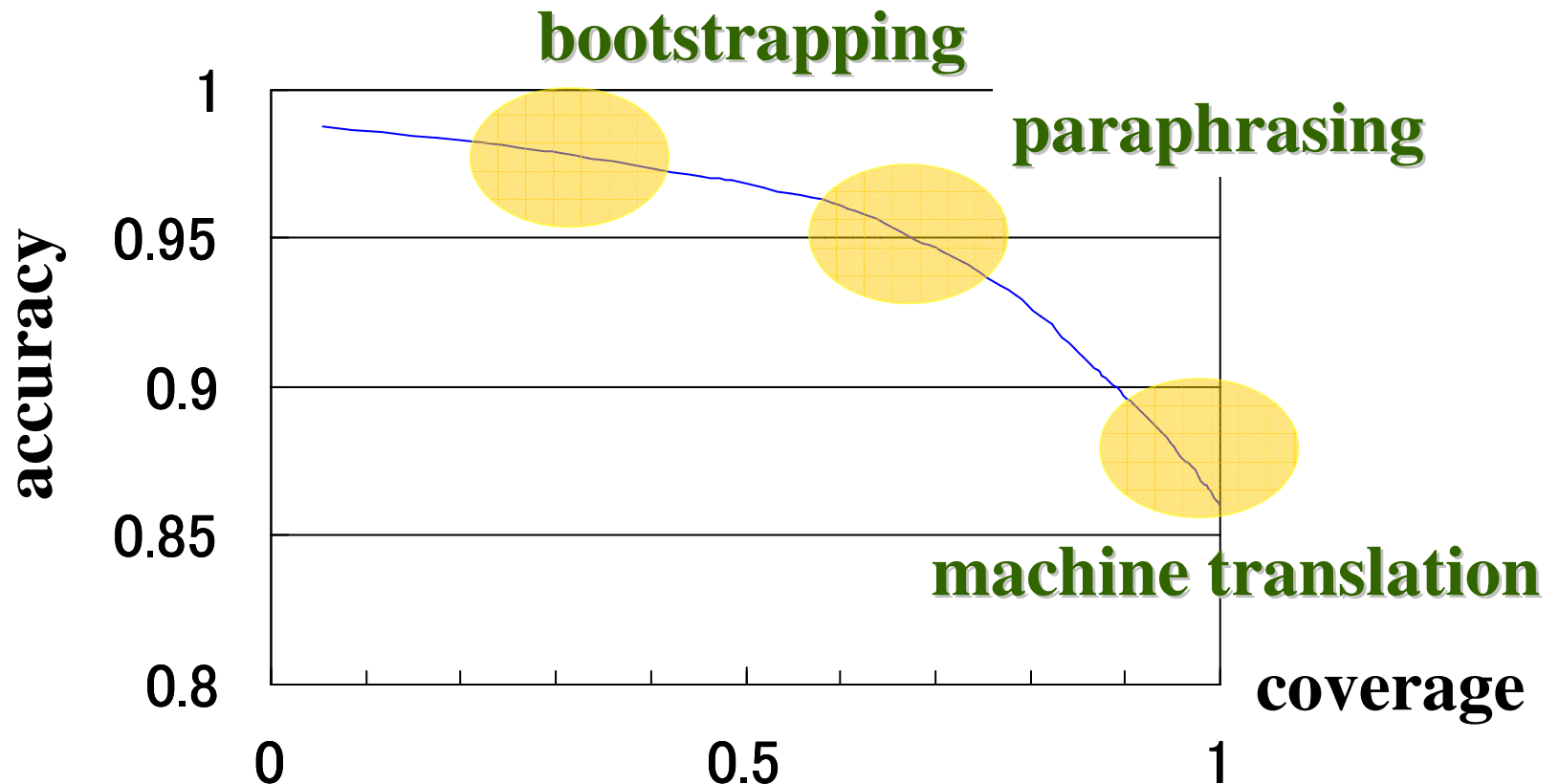
$$\text{coverage} = \frac{\text{\# of the decided relations}}{\text{\# of all the relations in the test set}}$$
$$\text{accuracy} = \frac{\text{\# of the correctly decided relations}}{\text{\# of the decided relations}}$$



Advantages

Probabilistic partial parsing

- ☐ The user can make a fine-grained arbitrary choice on the **trade-off** between coverage and accuracy
- ☐ Such trade-off choice makes the existing parsers of **wider application**



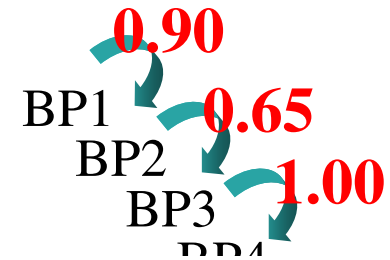
Estimation of DPs

Probabilistic partial parsing

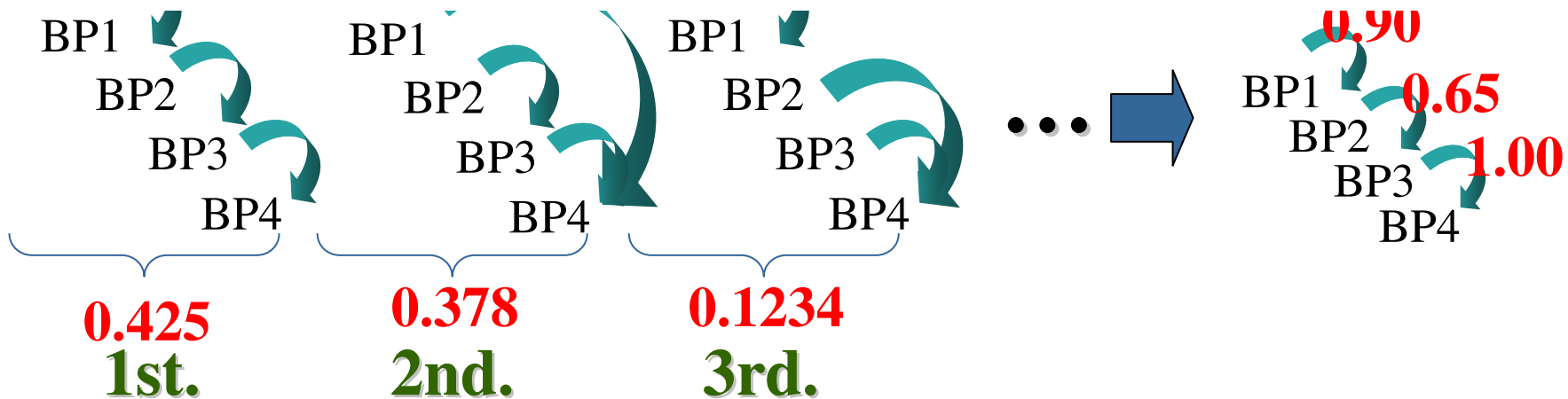
□ **Bottom-up models** (Collins, 1996) ,(Uchimoto et al., 1999)

❖ **Directly estimate DPs**

□ **Top-down models**



You can estimate DPs, whether you have a top-down model or a bottom-up model



**Committee-based
Probabilistic partial parsing**

Overview

Committee-based Probabilistic partial parsing

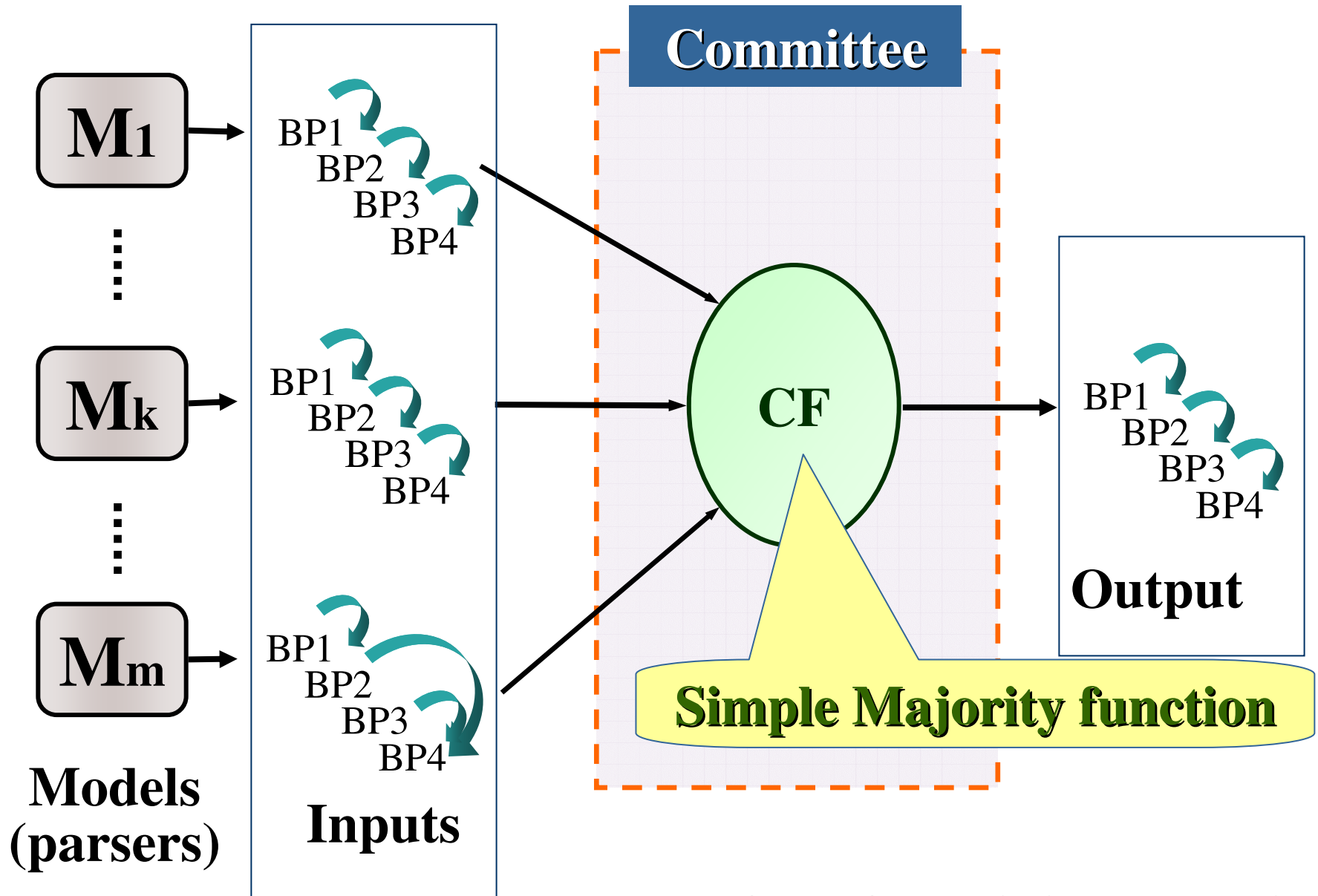
□ **Committee-based decision making is to combine the outputs from several different systems (e.g. parser) to make a better decision.**

- **POS tagging** (Halteren et al., 1998; Brill et al., 1999)
- **Parsing** (Henderson and Brill, 1999)
- **Word sense disambiguation** (Pedersen, 2000)
- **Machine translation** (Frederking and Nirenburg, 1994)
- **Speech recognition** (Fiscus, 1997)

□ **These works empirically demonstrated that combining different systems often achieved significant improvements over the previous best system.**

A basic scheme

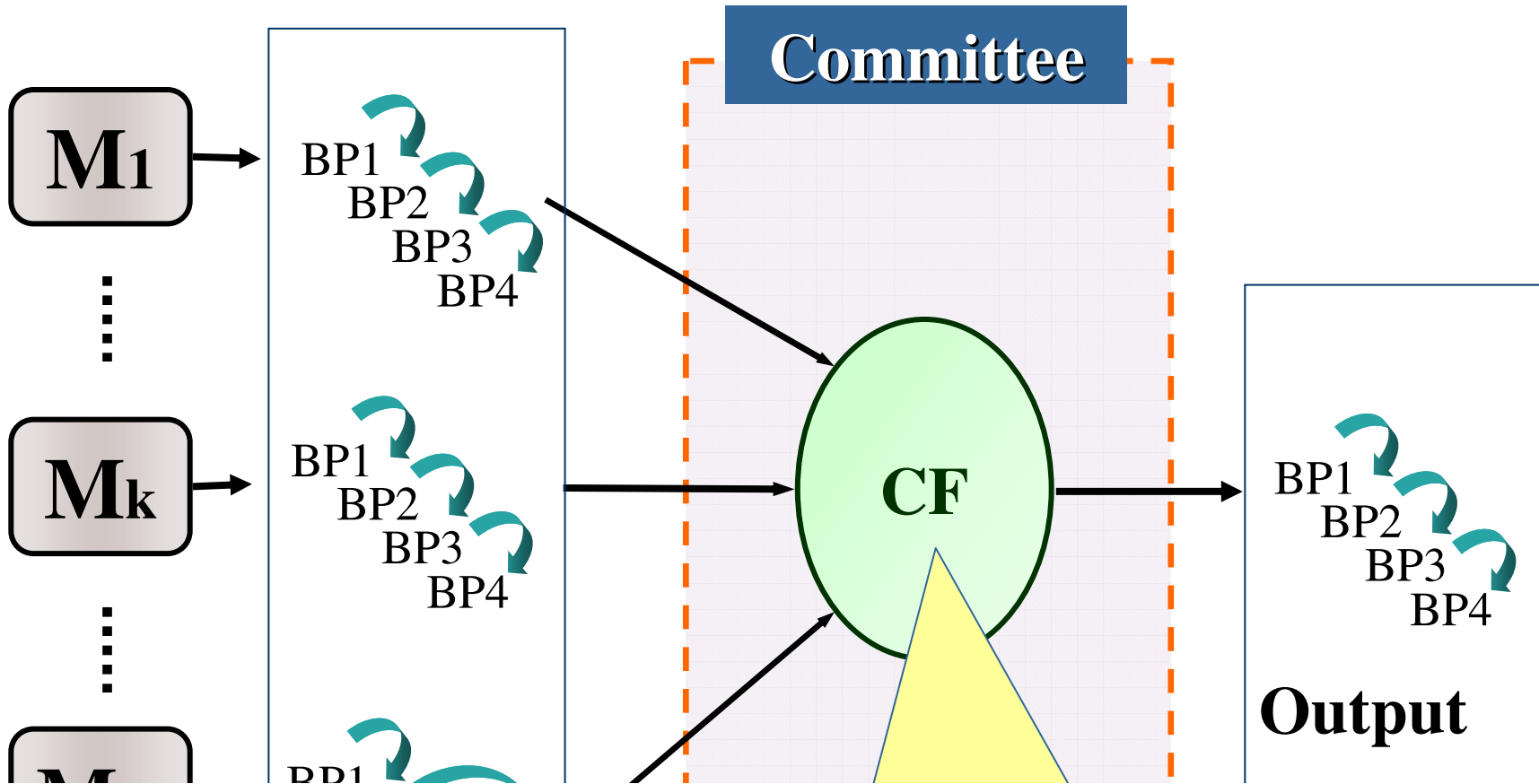
Committee-based Probabilistic partial parsing



CF : Combining Function

A basic scheme

Committee-based Probabilistic partial parsing



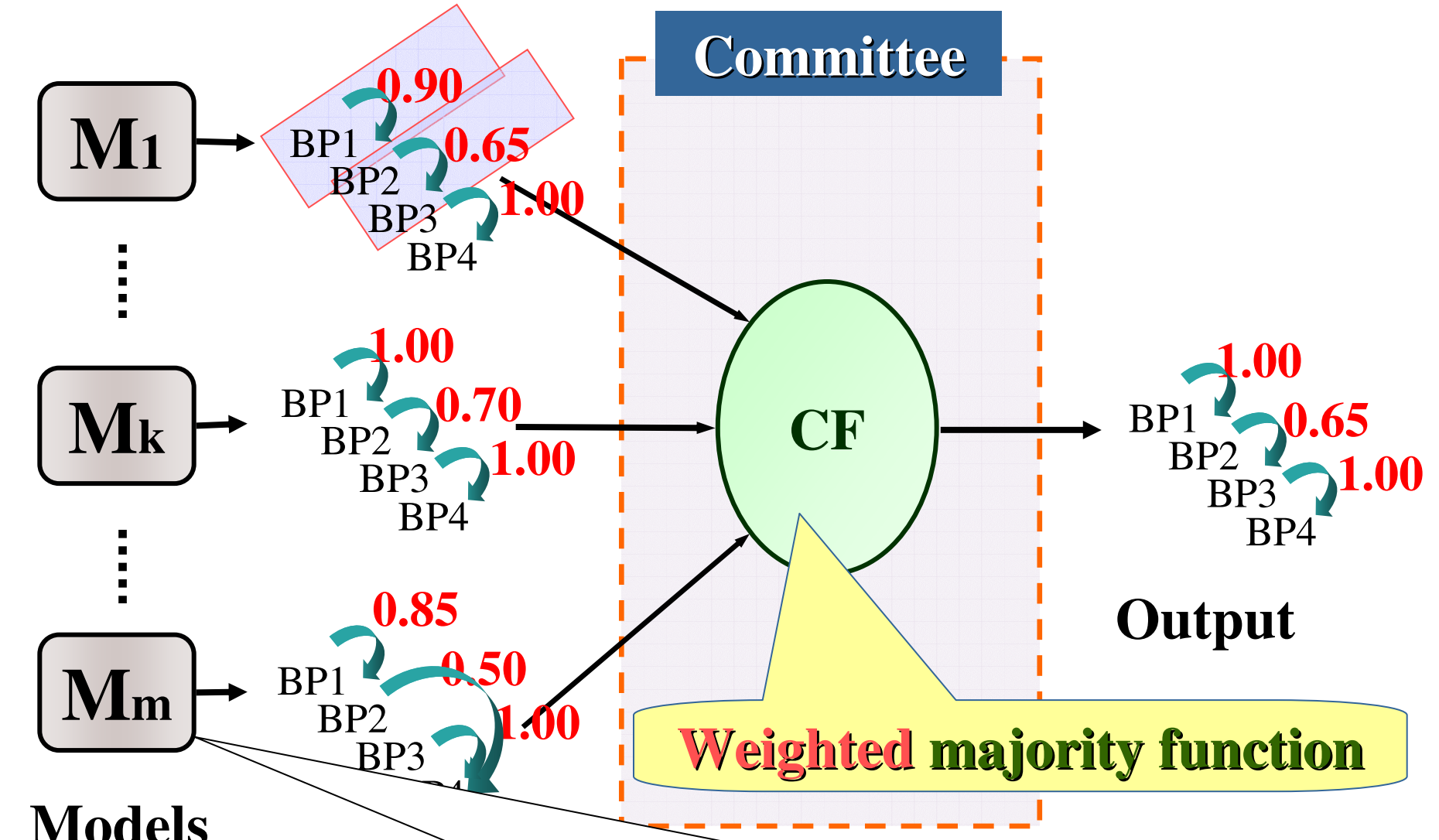
To realize partial parsing on this scheme,

⇒

the committee

would need to accept probabilistically annotated votes

Extension (1) : Probabilistic voting Committee-based Probabilistic partial parsing



Models
(parsers)

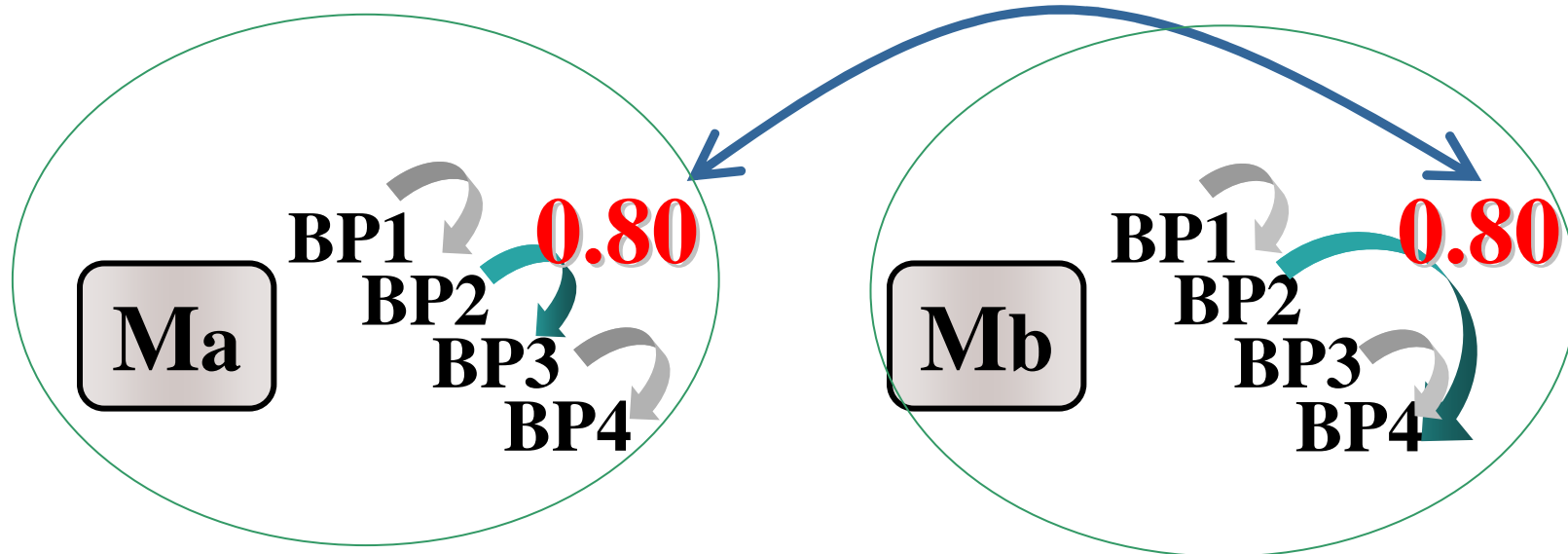
Most statistical parsers can be members of such a probabilistic voting committee

Extension (2) : Standardization

Committee-based
Probabilistic partial parsing

□ Reliability of dependency probabilities(DPs)

equally reliable?

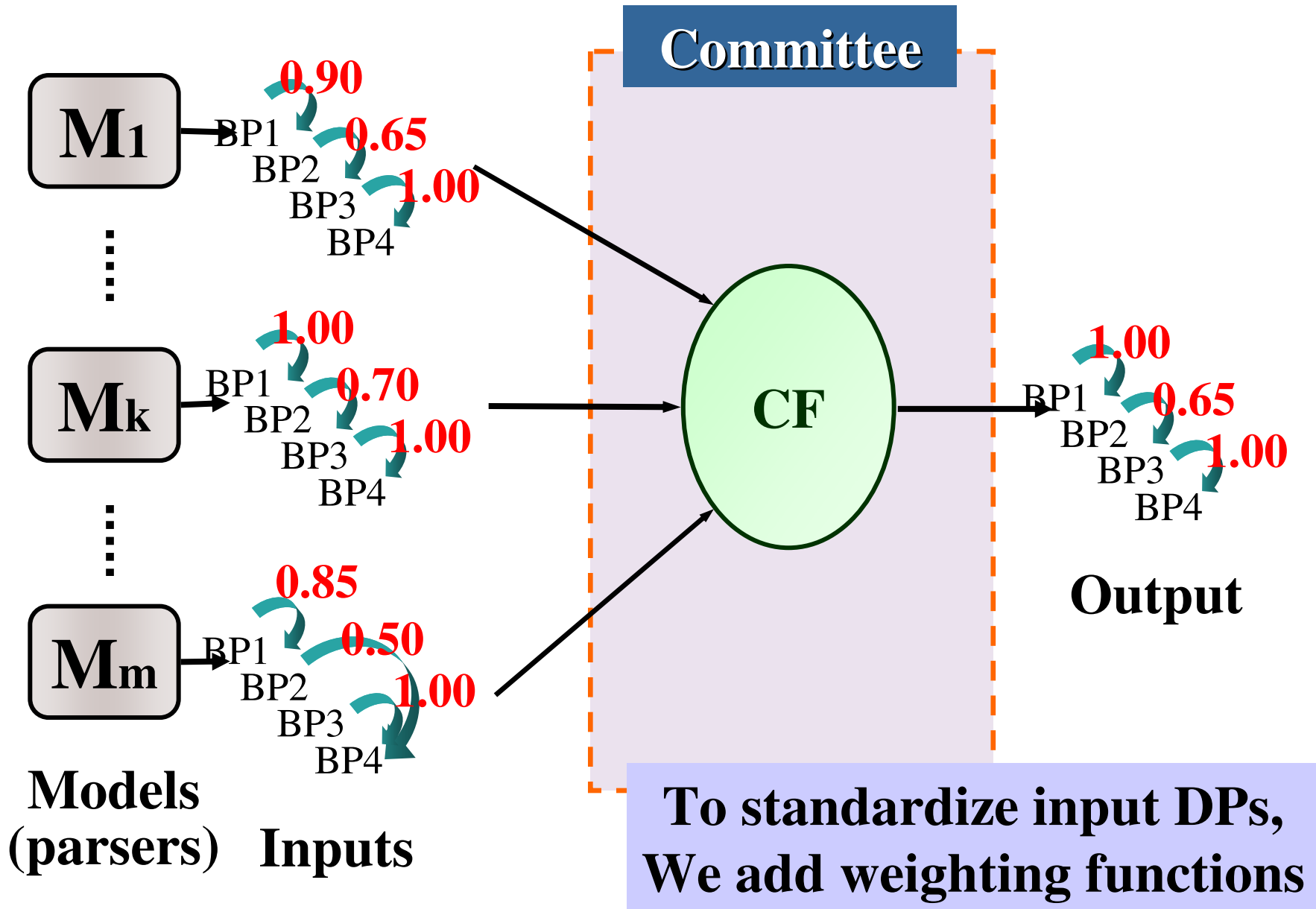


□ Reliability of DPs may differ depending on parsers

Standardization of DPs

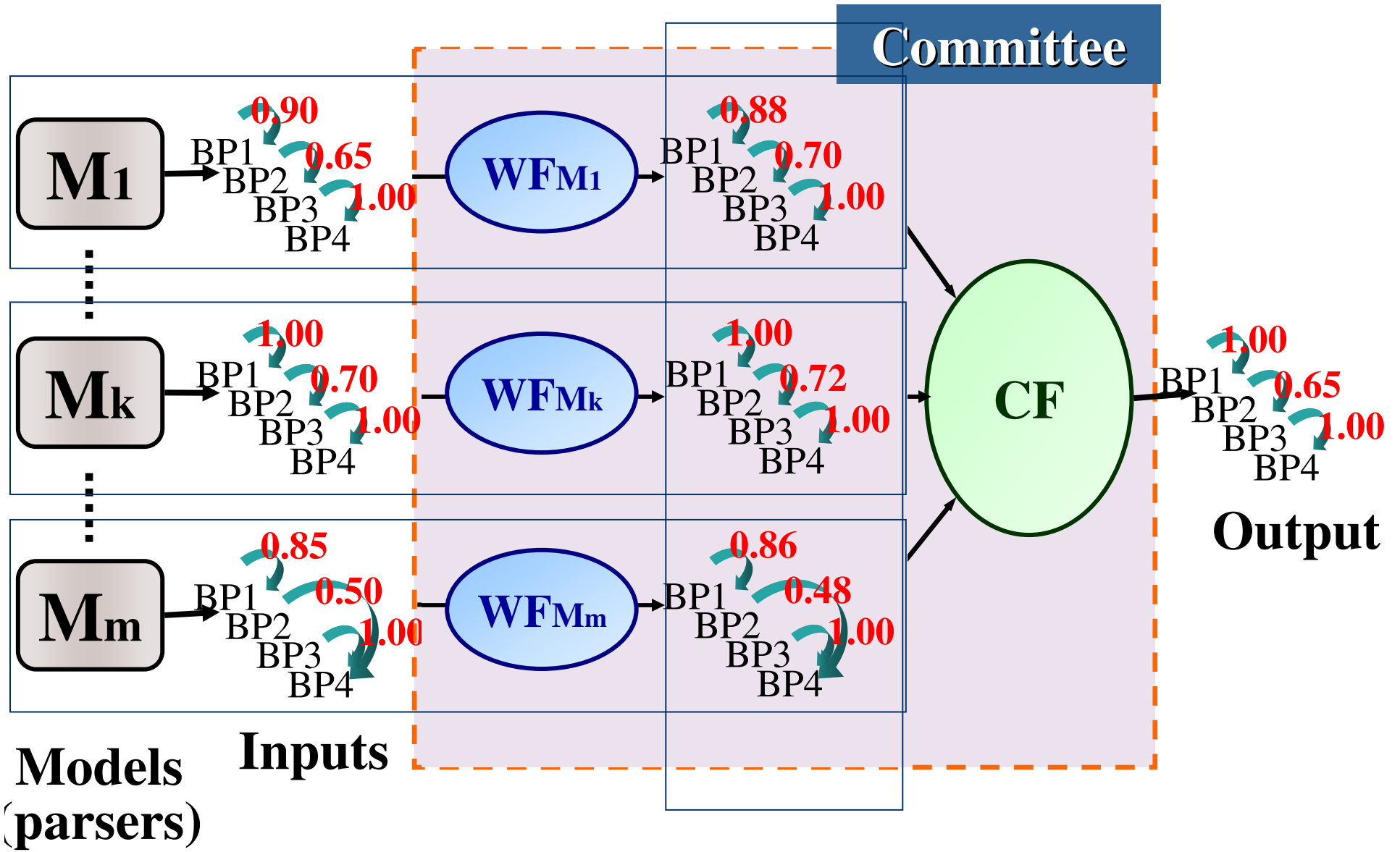
Extension (2) : Standardization

Committee-based Probabilistic partial parsing



Extension (2) : Standardization

Committee-based Probabilistic partial parsing

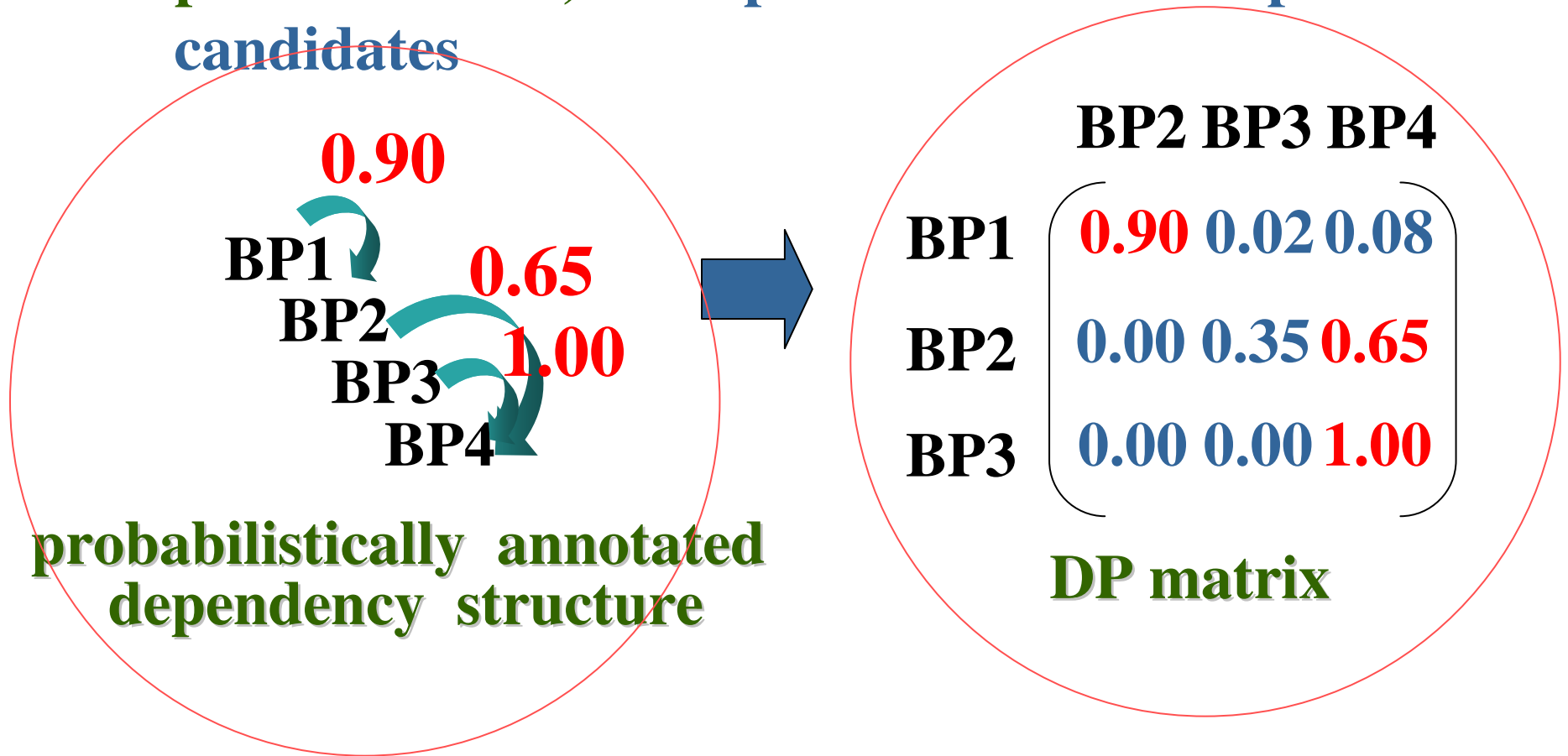


WF : Weighting Function

Extension (3) : Multiple voting

Committee-based Probabilistic partial parsing

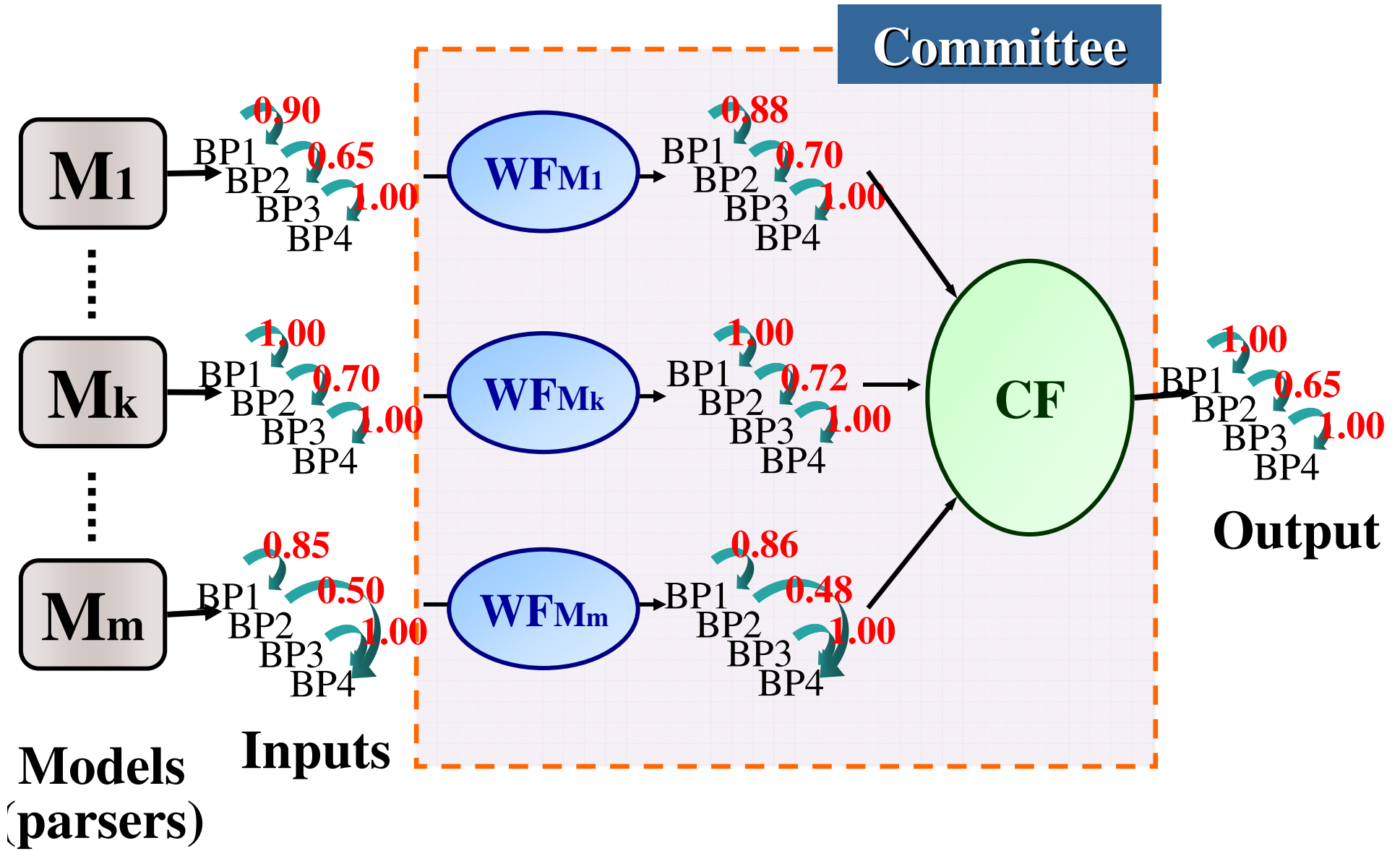
- Each member is allowed to cast (**probabilistically parameterized**) multiple votes for all the potential candidates



DP: Dependency Probability

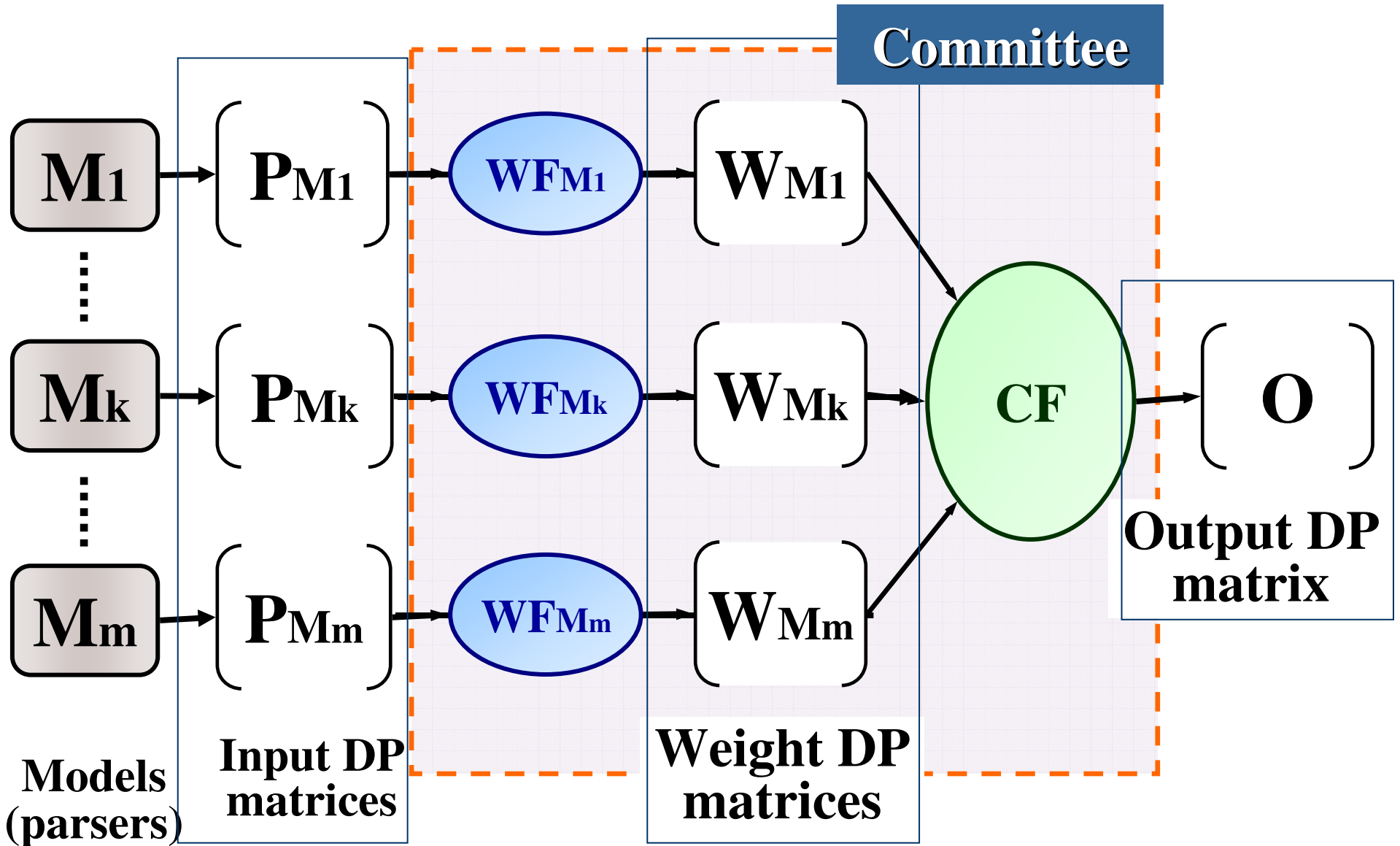
Extension (3) : Multiple voting

Committee-based
Probabilistic partial parsing



Extension (3) : Multiple voting

Committee-based Probabilistic partial parsing

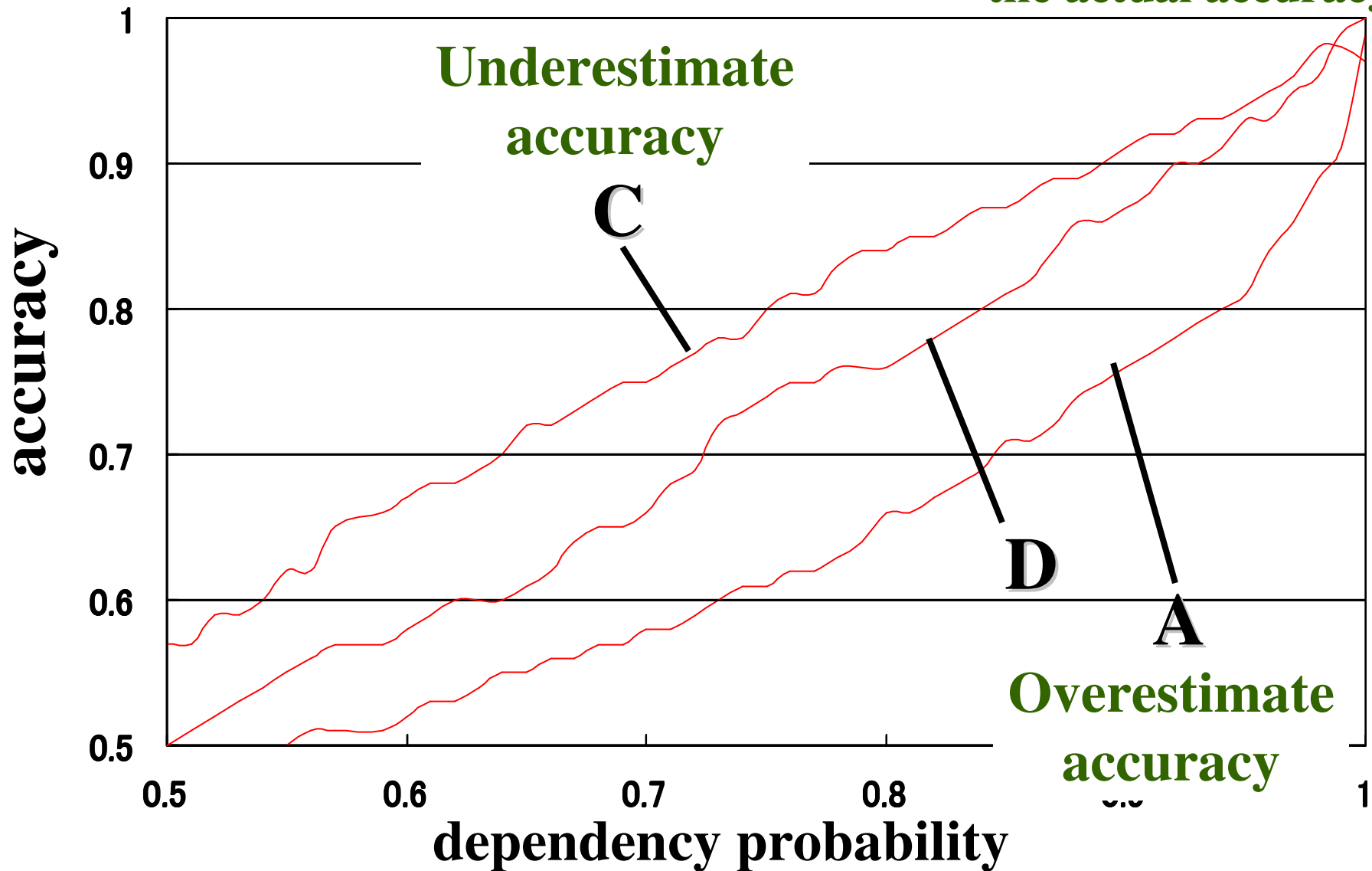


Generalized Committee-based Probabilistic Partial Parsing

Weighting functions

Committee-based Probabilistic partial parsing

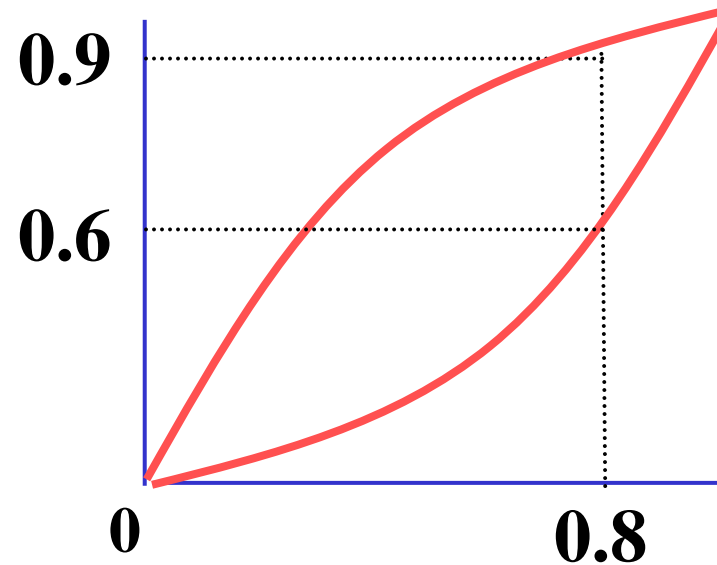
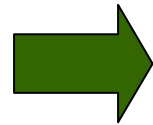
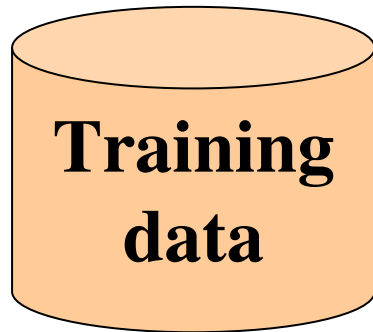
- ❑ A bare DP may not a precise estimation of the actual accuracy



Weighting functions

Committee-based Probabilistic partial parsing

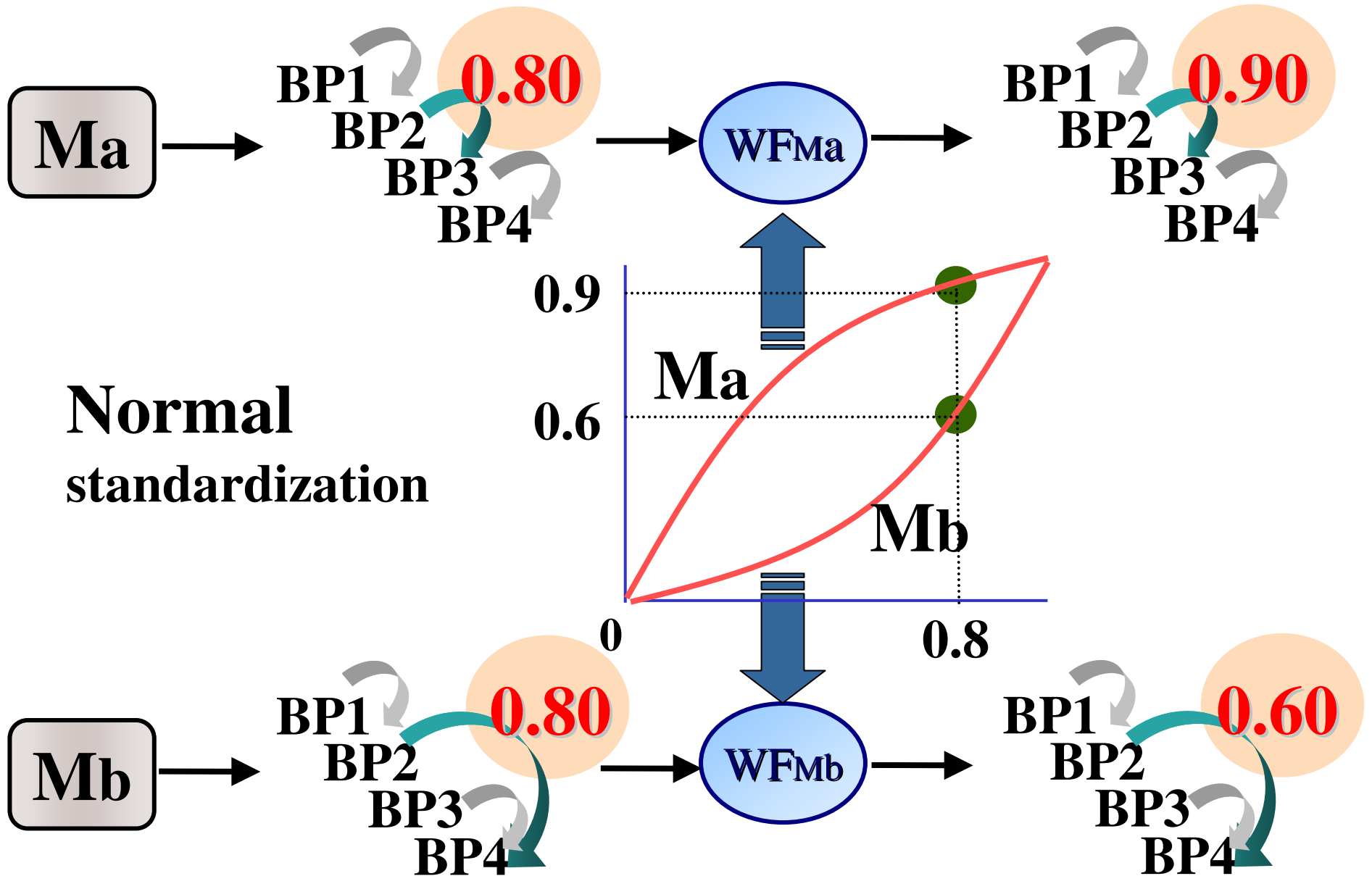
You can standardize input DPs by referring to P-A curves acquired from some training data



P-A curves

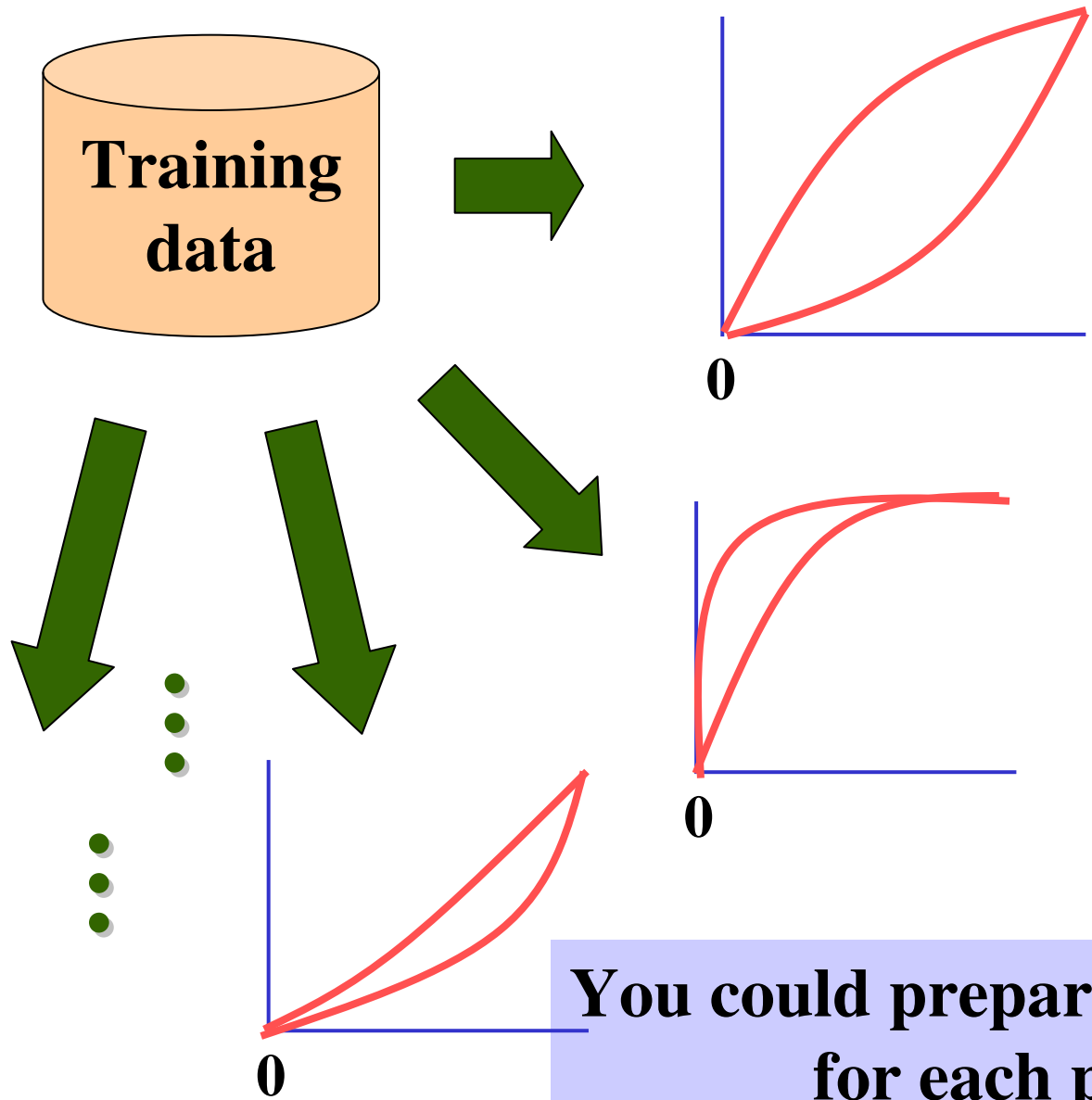
Weighting functions : Normal

Committee-based Probabilistic partial parsing



Weighting functions : Class

Committee-based Probabilistic partial parsing



**Advarbial
dep. relations**

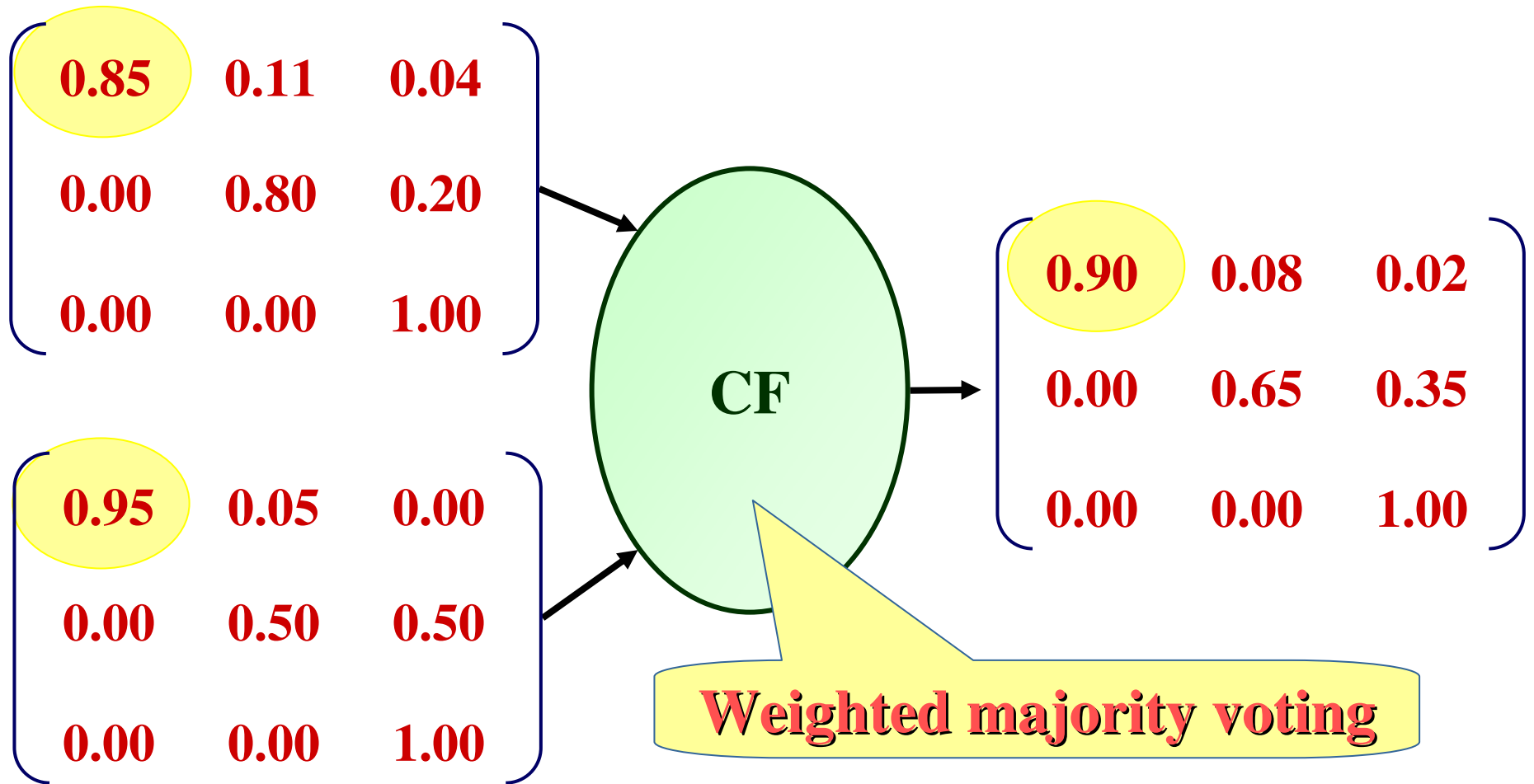
**Class-based
standardization**

**Adnominal
dep. relations**

**You could prepare weighting functions
for each problem class**

Combining function

Committee-based Probabilistic partial parsing



☐ Averaging of the given weight matrices

Summary

Committee-based Probabilistic partial parsing

Our committee-based scheme:

- (a) accepts **probabilistic parameterized votes** as its input
- (b) accepts **multiple voting**
- (c) considers the **standardization** of original input votes
- (d) outputs a **DP matrix** as a final decision

DP: Dependency Probability

Related works

Committee-based Probabilistic partial parsing

□ Our voting scheme =

Generalization of existing voting techniques for NLP:

- ❖ Probabilistic multiple voting
- ❖ Standardization
- ❖ DP matrix output (coverage/accuracy trade-off)

□ Previous voting techniques

- POS tagging (Halteren et al., 1998)
- Parsing (Henderson and Brill, 1999)
- . . .

**Not accept
multiple voting**

**Not accept
probabilistic voting**

Experiments

committee members (parsers)

Experiments

KANA

(Ehara, 1998) : a bottom-up model based on maximum entropy estimation

CHAGAKE

(Fujio et al., 1998) : an extension of the bottom-up model proposed Collins (Collins, 1996)

Kanayama's parser

(Kanayama et al., 1999) : a bottom-up model coupled with a HPSG

Shirai's parser

(Shirai et al., 1998) : a top-down model incorporating lexical collocation statistics

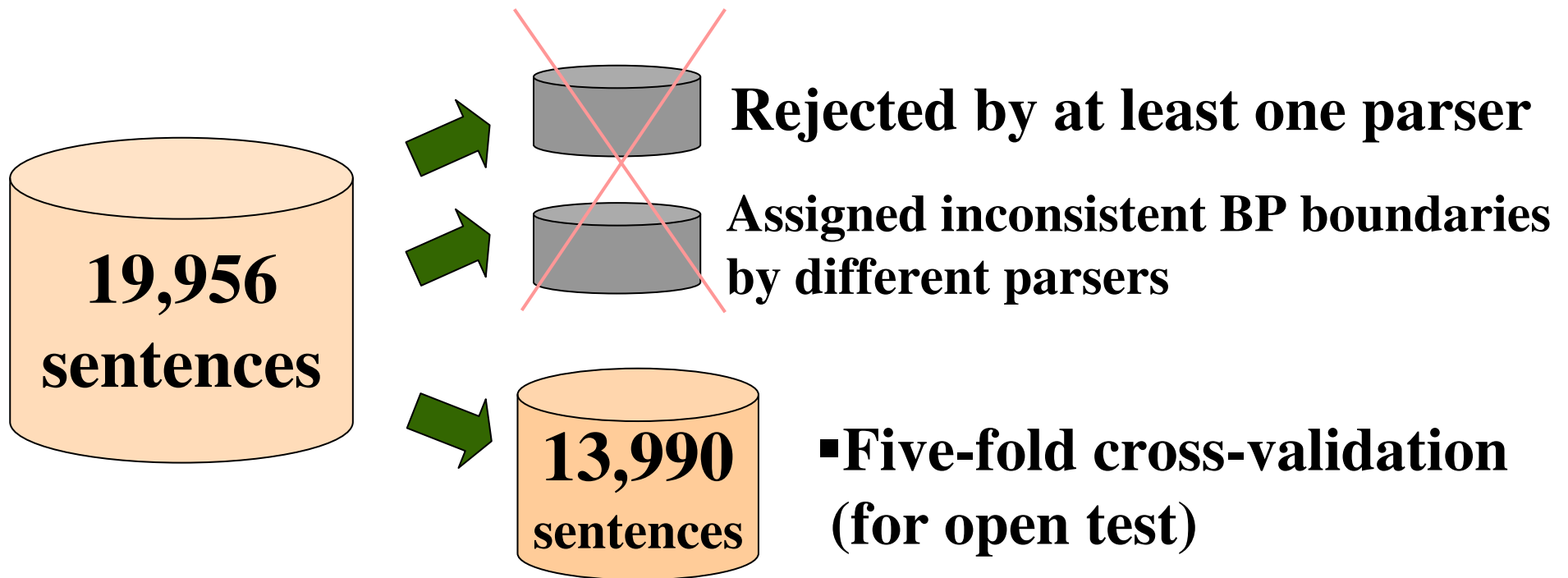
Peach Pie Parser

(Uchimoto et al., 1999) : a bottom up model based on maximum entropy estimation

Training / test sets

Experiments

- ❑ **Kyoto corpus(ver2.0)** (Kurohashi et al., 1997)
 - ❖ collection of Japanese newspaper articles
 - ❖ annotated in terms of :



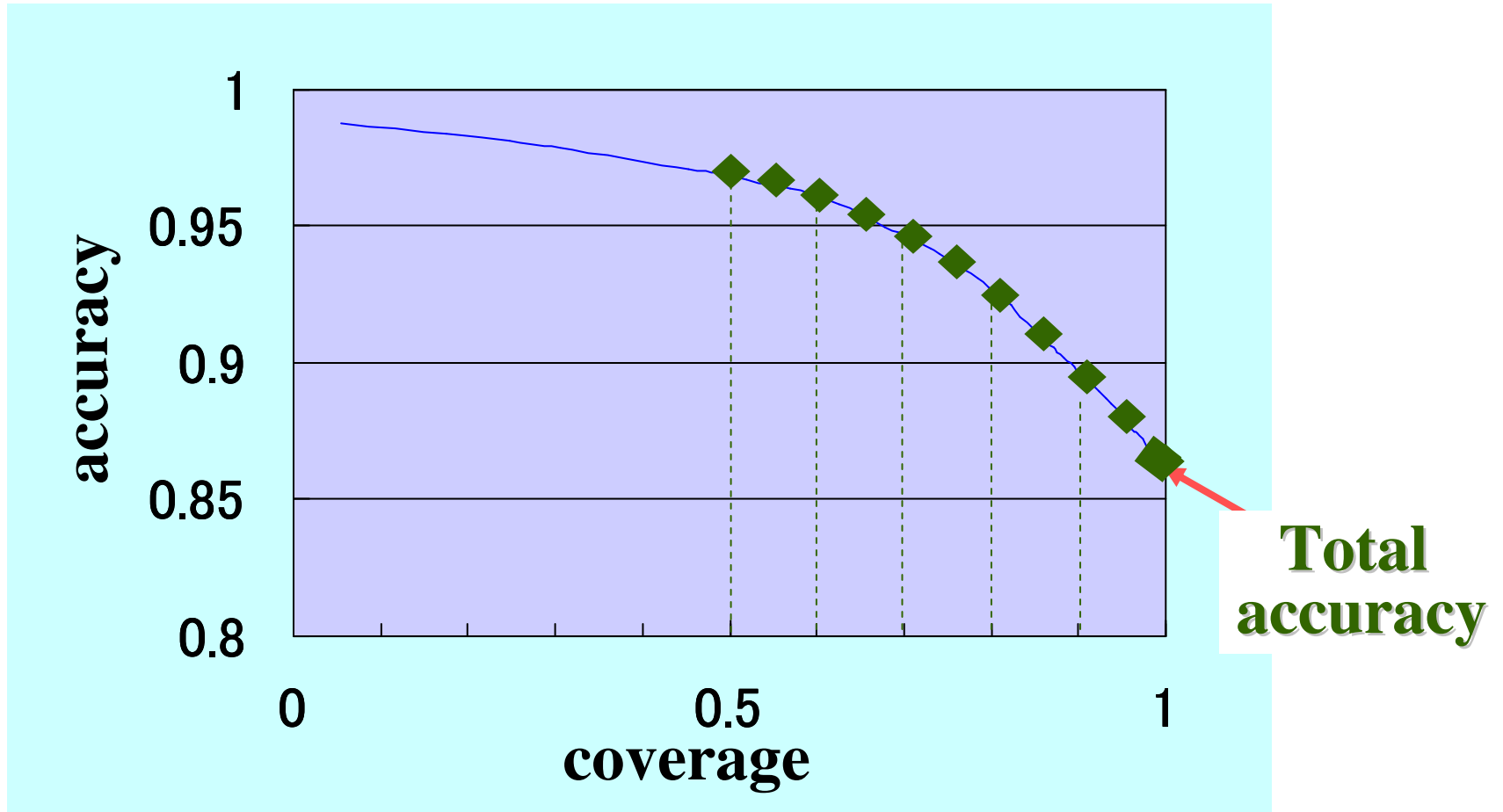
Performance of each individual model Experiments

Model (parser)	Total accuracy	11-point accuracy
A	0.8974	0.9607
B	0.8551	0.9281
C	0.8586	0.9291
D	0.8470	0.9266
E	0.7885	0.8567

- **Total accuracy and 11-point accuracy are both given by C-A curve**

C-A curve

Experiments



- ❑ 11-point accuracy is a summary of the C-A curve, which is given by the average of the accuracy of 11 points

Accuracy of each individual model

Experiments

Model (parser)	Total accuracy	11-point accuracy	
A	0.8974	0.9607	← Optimal
B	0.8551	0.9281	} Sub-optimal Comparable
C	0.8586	0.9291	
D	0.8470	0.9266	
E	0.7885	0.8567	

□ **Model A is significantly better than other models**

Issue (1) : Probabilistic voting

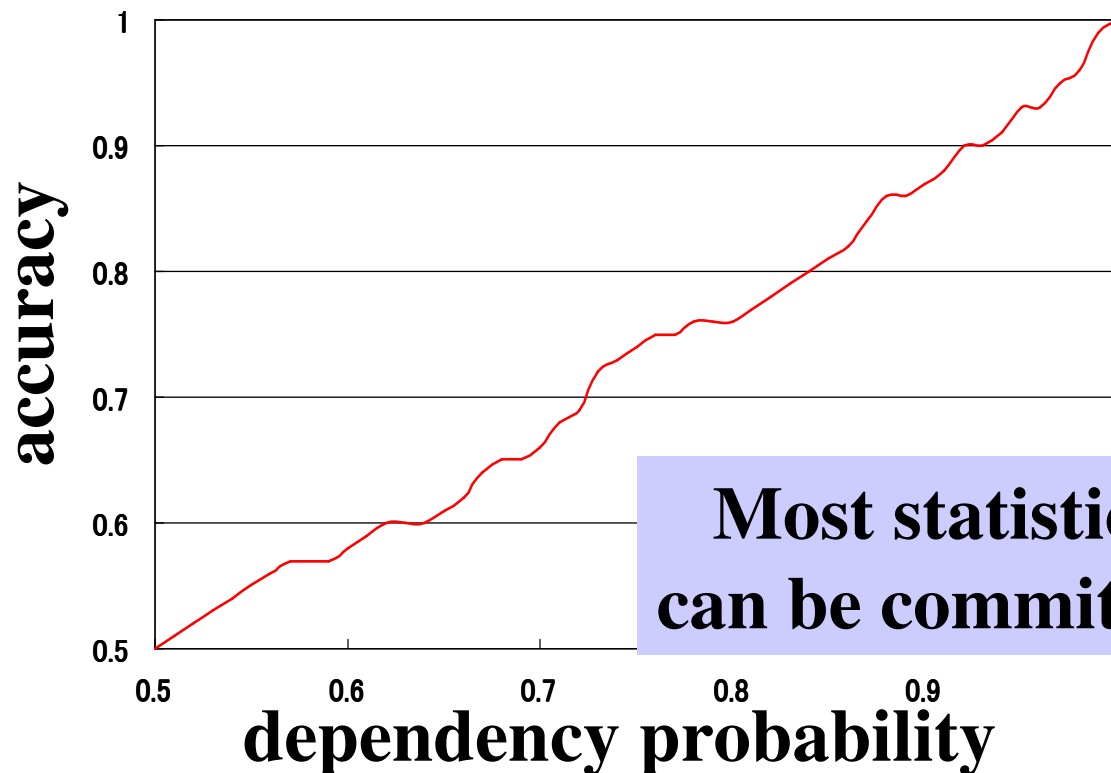
Experiments

❑ **Can we easily gather committee members?**

Yes!

❖ **Shirai's parser**(Shirai et al., 1998) :

- a top-down model (not provide DPs directly)
- By using n-best dependency structure candidates, we were able to estimate DPs reasonably correctly



**Most statistical parsers
can be committee members**

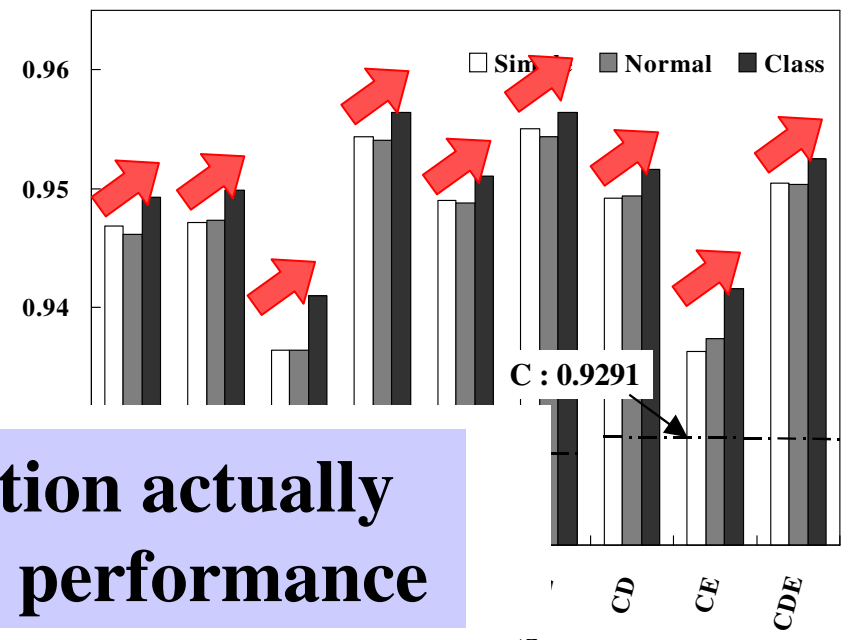
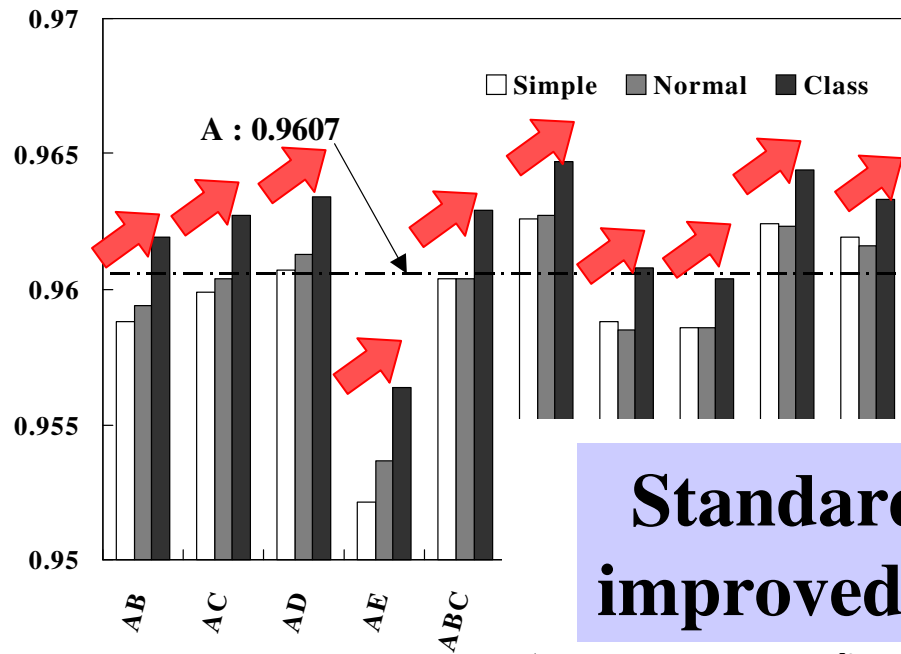
Issue (2) : Standardization

Experiments

Is standardization actually effective?

Yes!

11-point accuracy



Standardization actually improved the performance

A included

B included

C included

committee

Issue (3) : Multiple voting

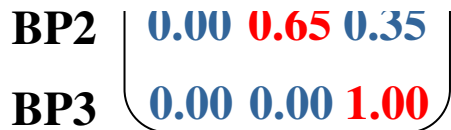
Experiments

Does multiple voting improve the performance?

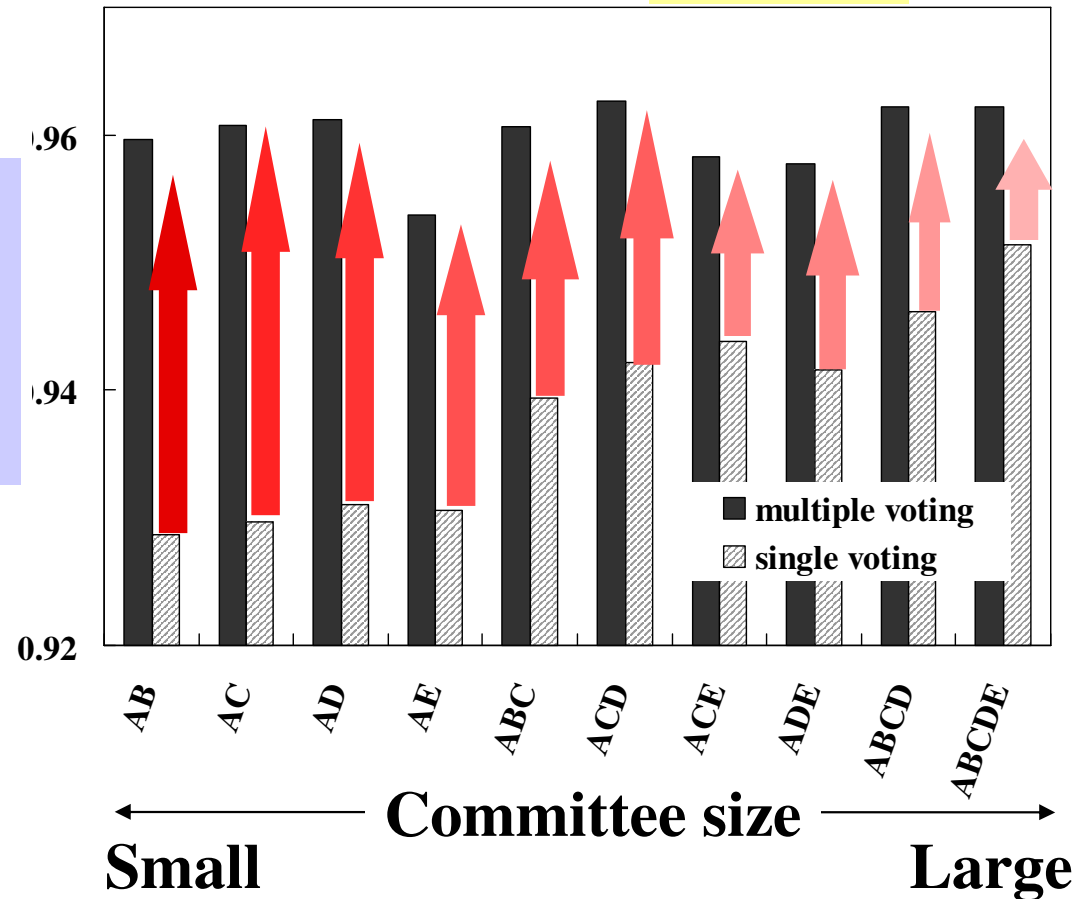
Yes!



At least when the size of a committee is small, multiple voting significantly outperformed single voting



Multiple voting



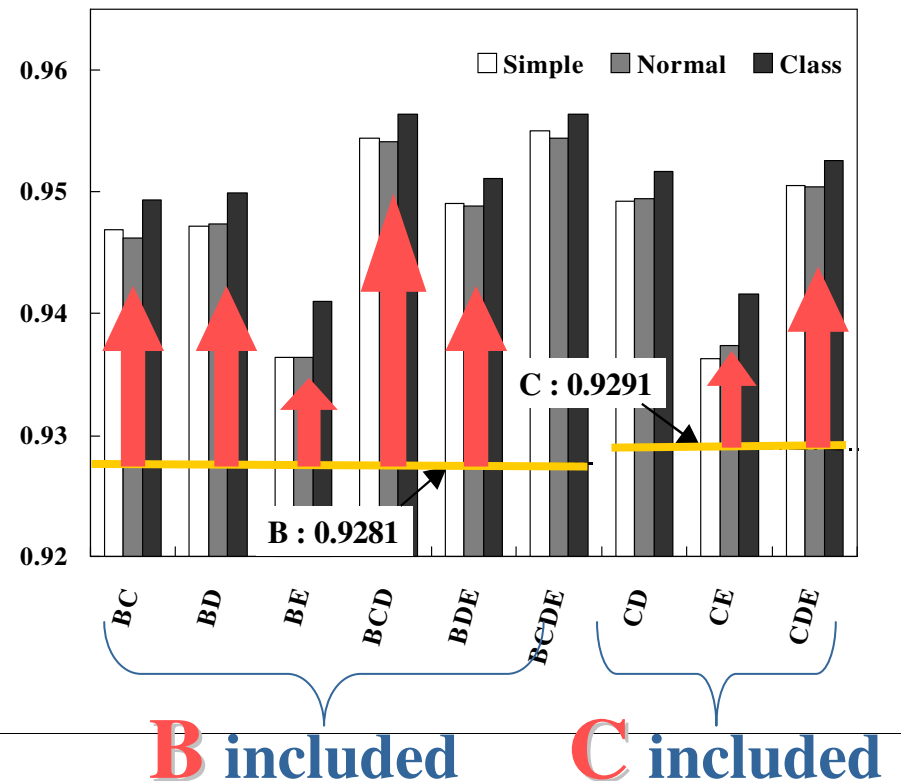
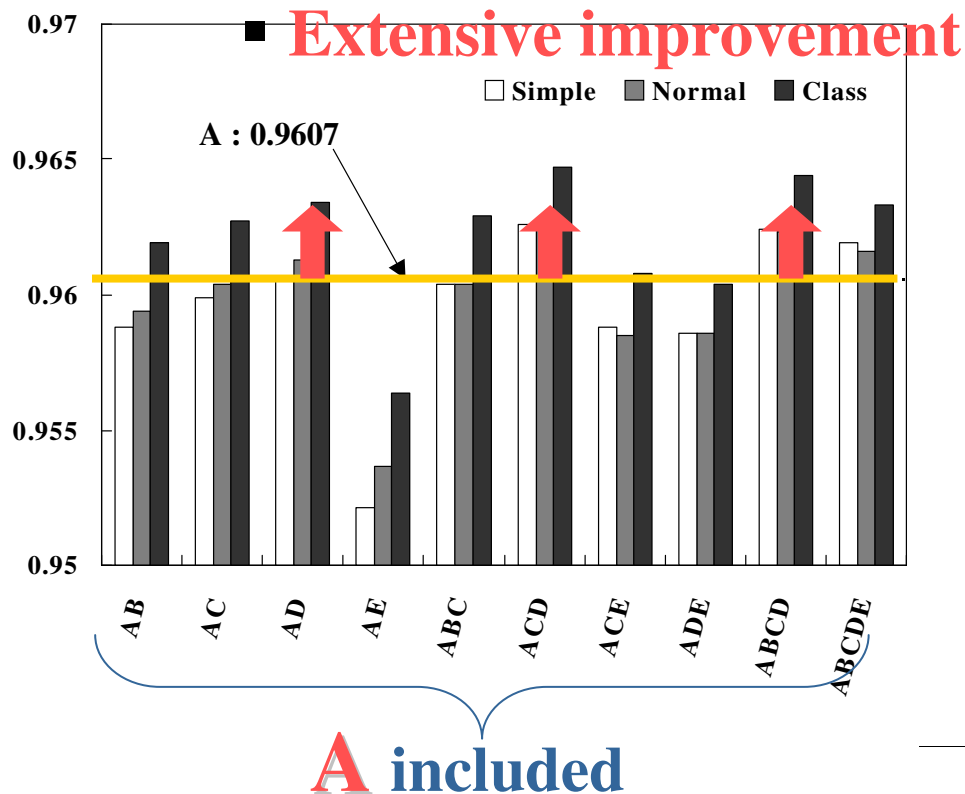
Issue (4) : Contributions of a committee Experiments

❑ Does combining parsers actually improve

❖ Including the optimal model A, the performance?

▪ Not very visible improvement.

❖ Including the comparable members such as BC or BD



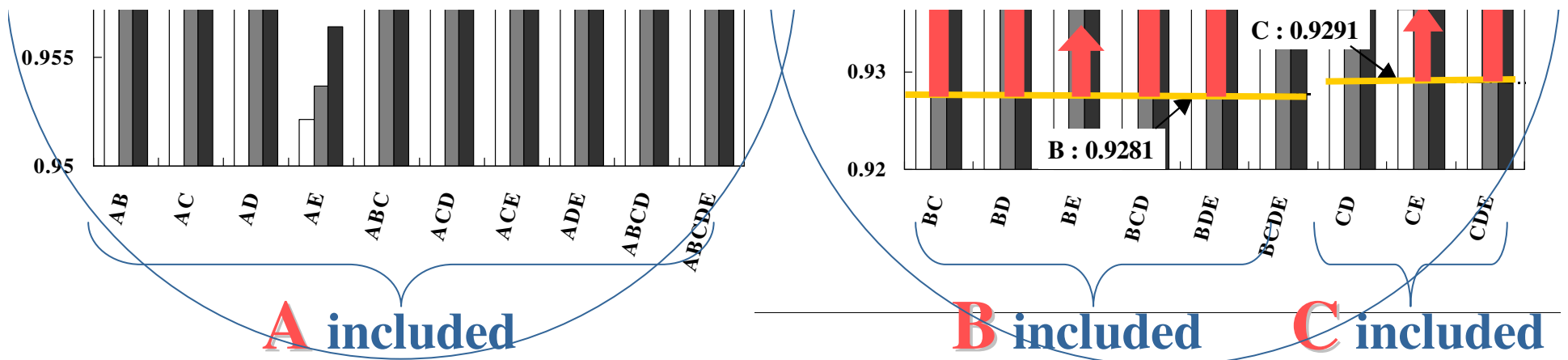
Issue (4) : Contributions of a committee Experiments

□ **Does combining parsers actually improve**

❖ **Including the optimal model A, the performance?**

▪ **Not very visible improvement.**

If we have another optimal parser that was comparable to parser A, then we might achieve significant improvements even in case where parser A participates



Conclusion

- We proposed a general committee-based framework that can be coupled with probabilistic partial parsing

Findings through experiments

- (a) Both **multiple voting** and **vote standardization** effectively work in committee-based partial parsing
- (b) If more than two comparably competent optimal models are available, it is likely to be worthwhile to combine them
- (c) Our scheme also enables a non-parametric rule-based parser to make a good contribution