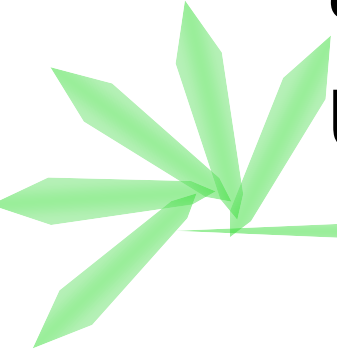
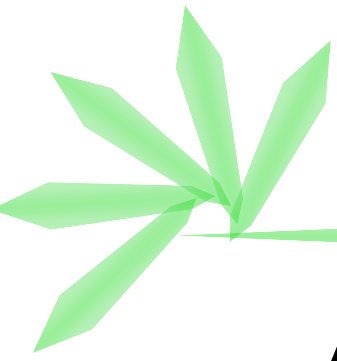




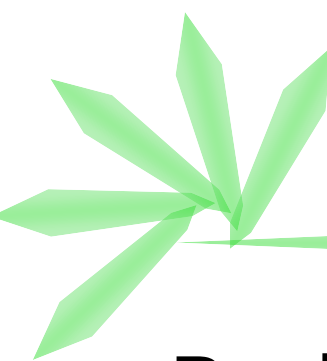
Automatic Construction of a Bilingual Thesaurus using Citation Analysis

2011/10/24

Hidetsugu Nanba, Saori Mayumi, Toshiyuki Takezawa

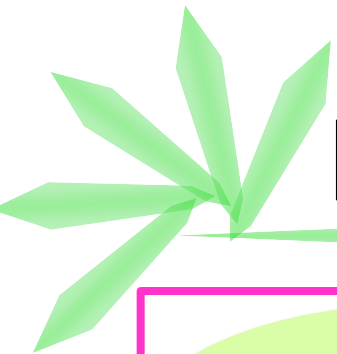
Hiroshima City University



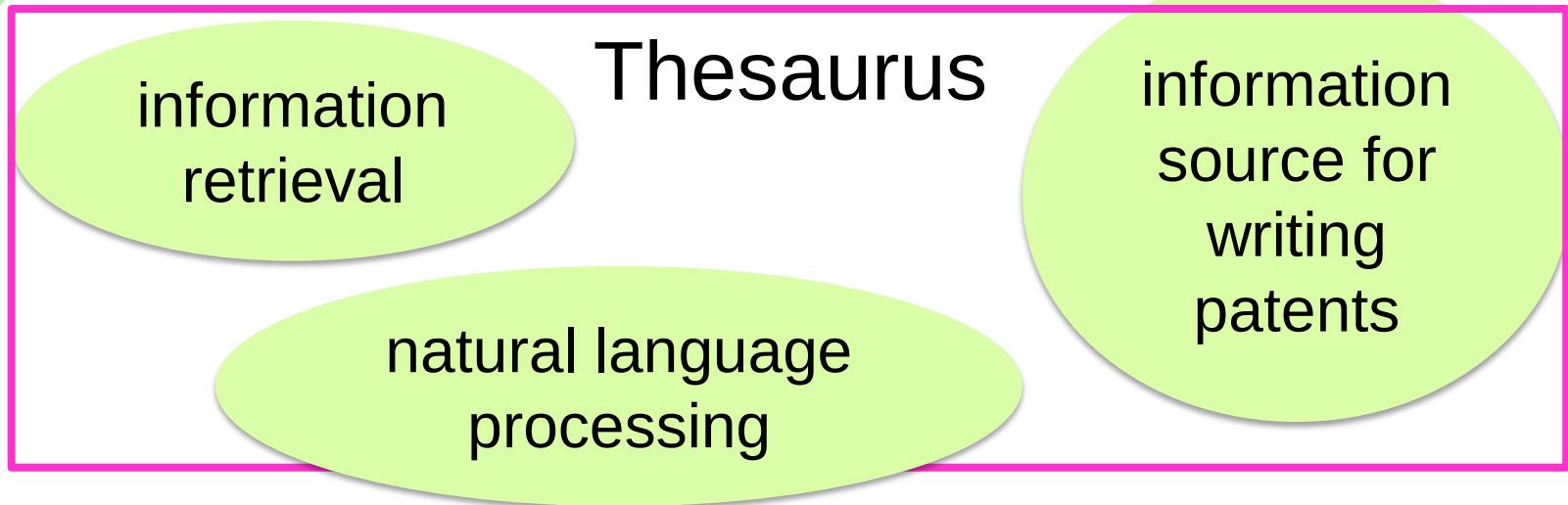


Contents

- Background and Purpose
- Hyponymy Extraction (Related Work)
- Automatic Alignment between English and Japanese Hypernym-Hyponym Pairs
- Experiments
- Conclusion



Background and Purpose



costly to construct and maintain manually

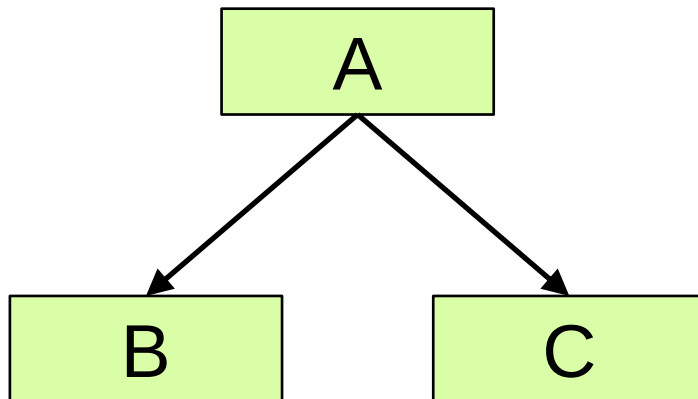


automatic construction of a thesaurus

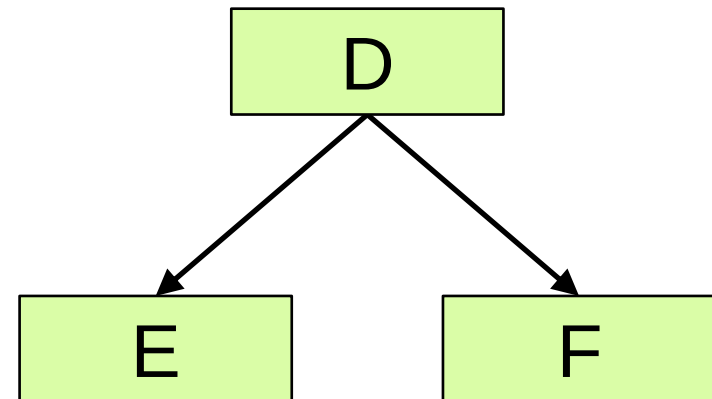
Hyponymy Extraction

- Pattern-based approach [Hearst 1992]
- Apply Hearst's method to patent docs [Nanba 2007]

A such as B, C



EやFなどのD



165.242.101.30/cgi-bin/terms_e.cgi?keywords=cutting+tool

Google

これは 英語 のページです。翻訳しますか? 翻訳 いいえ 英語を翻訳しない オプション

[\[Top\]](#)

cutting tool

(synonym)

(hyponym)

[drill\(50\)](#) [end mill\(39\)](#) [pilot mill\(20\)](#) [router bite\(17\)](#) [face milling cutter\(16\)](#)
[knife\(16\)](#) [scissors\(16\)](#) [milling cutter\(11\)](#) [anchor\(10\)](#) [drill bit\(10\)](#) [drill](#)
[bite\(10\)](#) [milling guide\(10\)](#) [Thompson blade\(9\)](#) [cutter\(9\)](#) [router\(9\)](#)
[reamers\(8\)](#) [see\(8\)](#) [cutting tool\(7\)](#) [file\(7\)](#) [reamer\(7\)](#) [rotary cutter\(7\)](#)
[diamond cutter\(6\)](#) [housing\(5\)](#) [jig\(5\)](#) [ball-end mill\(4\)](#) [conventional](#)
[router\(4\)](#) [crown\(4\)](#) [laser\(4\)](#) [lathe\(4\)](#) [razor\(4\)](#) [saw blade\(4\)](#) [tip\(4\)](#) [tool\(4\)](#)
[turning tool\(4\)](#) [utility knife\(4\)](#) [contour cutter\(3\)](#) [die grinder\(3\)](#) [drilling\(3\)](#)
[hand-held drill\(3\)](#) [knives\(3\)](#) [pair\(3\)](#) [press\(3\)](#) [pullulan film\(3\)](#) [punch\(3\)](#)
[razor blade\(3\)](#) [saw\(3\)](#) [scalpel\(3\)](#) [shear\(3\)](#) [torch\(3\)](#) [twist drill\(3\)](#) [type\(3\)](#)
[bite\(2\)](#) [blade\(2\)](#) [boring tool\(2\)](#) [broach\(2\)](#) [center drill\(2\)](#) [circular see\(2\)](#)
[diamond blade\(2\)](#) [diamond wheel\(2\)](#) [dicing see\(2\)](#) [disc\(2\)](#) [drill 460](#)
[relative\(2\)](#) [drilling tool\(2\)](#) [grinder wheel\(2\)](#) [grindstone\(2\)](#) [hand shear\(2\)](#)
[helical cutter\(2\)](#) [insert\(2\)](#) [kitchen knives\(2\)](#) [linear edge blade\(2\)](#) [lock](#)

[\[Top\]](#)

「切断手段」の検索結果をすっきり表示

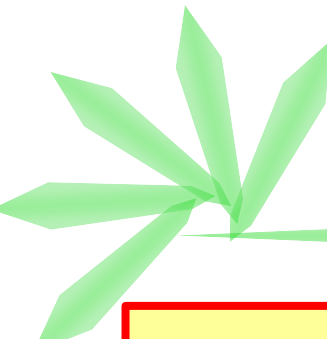
切断手段

下位語

カッター(124) 切断刃(20) ナイフ(17) レーザ(16) 切断(15) レーザー(14) 回転刃(12) 鋏(11) ミシン目(7) 旋盤(7) ウォータージェット(6) ハサミ(6) プレス(6) レーザ光(6) 刃物(5) カッターナ
イフ(4) パンチ(4) リッター(4) ロータリーカッター(4) 移動(4) 加熱線条(4) 鋸(4) 鋸刃(4) 刃(4)
クラッチ(3) コア(3) ダイヤモンドカッター(3) ヒータ(3) レーザーカット(3) ワイ

Cutter (124), Cutter blade (20), Knife (17),
Laser (16), Rotary blade (16), Scissors (11),
Perforation (7) ...

クリート(1) シール手段(1) シャー(1) ジェット水流(1) スライシング(1) スライシングマシン(1)
ダイヤモンドカッター刃(1) ディスクカッター(1) トムソン金型刃(1) バイト式(1) バンドリイ(1) ピ



Background and Purpose

Construct E/J patent thesaurus

Extract hyponymy from E/J patent documents

SMT technique

Align E-J terms

Citation analysis

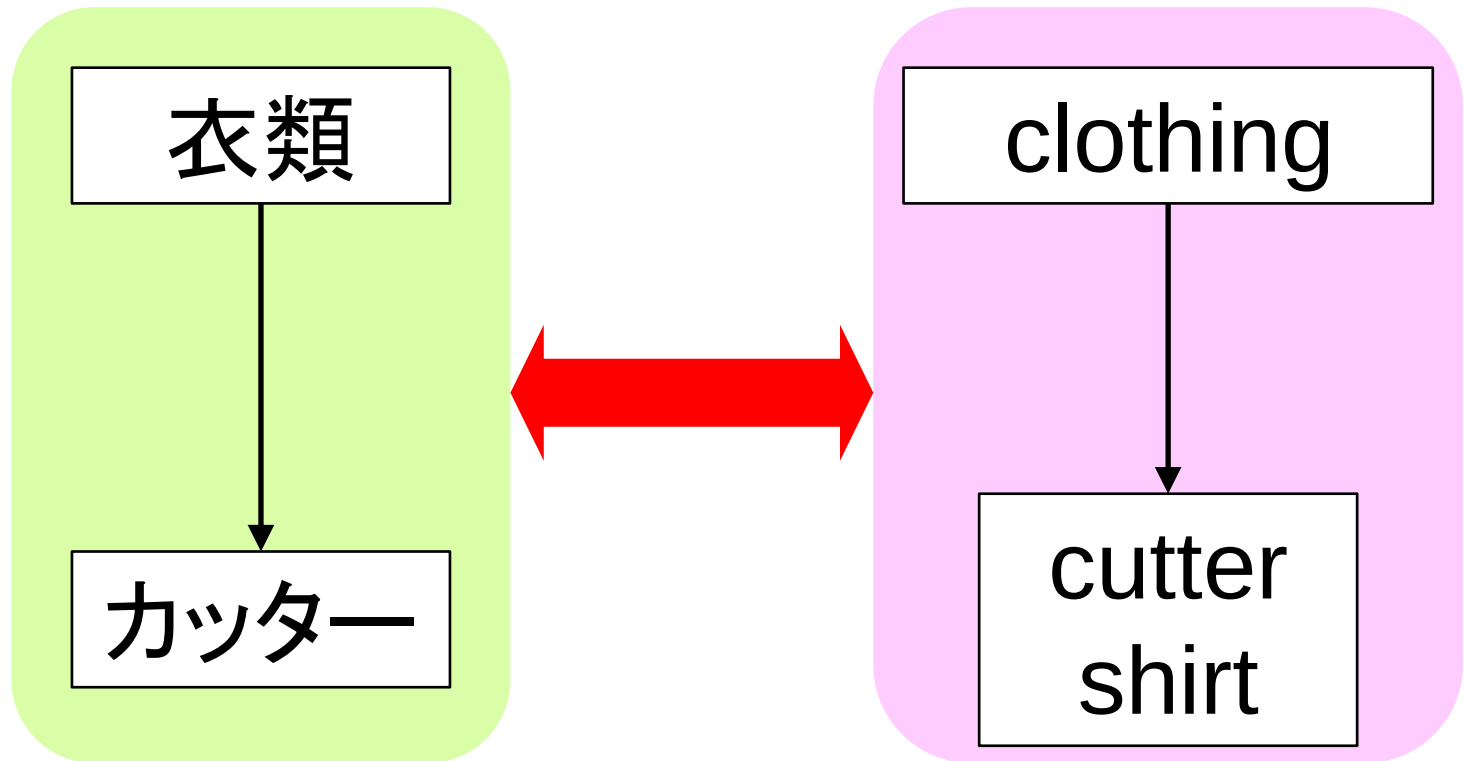
Bilingual thesaurus

Cross-lingual IR

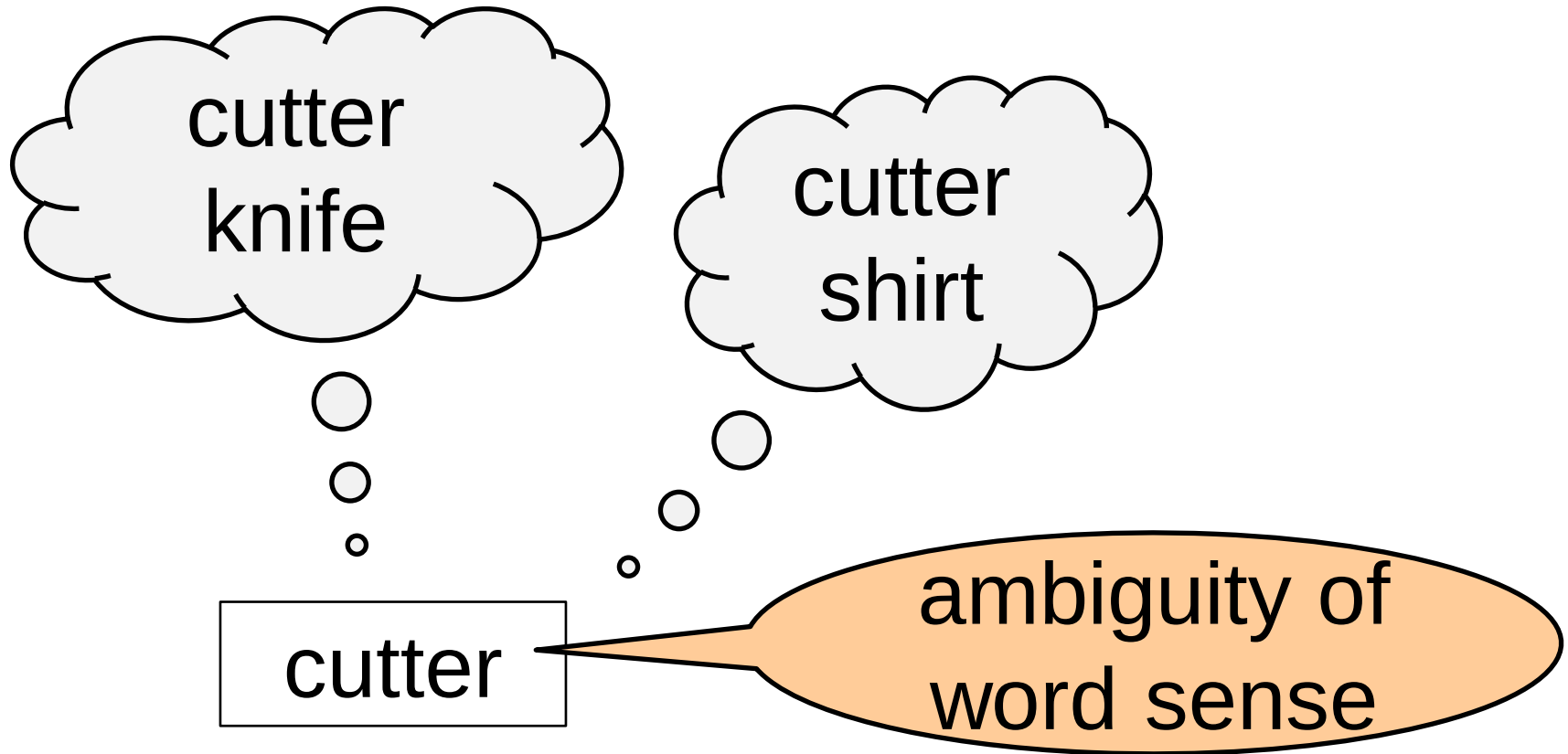
NLP

information source
for writing patents

Automatic Alignment between English and Japanese Hypernym- Hyponym Pairs

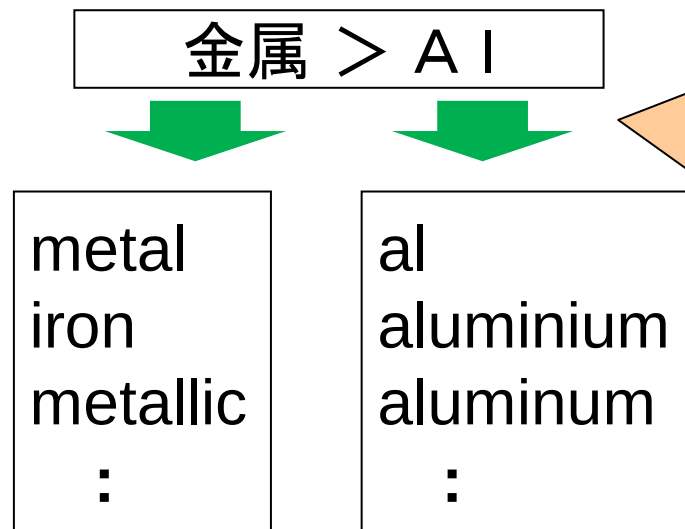


Automatic Alignment between English and Japanese Hypernym- Hyponym Pairs



Alignment Using a Phrase Table

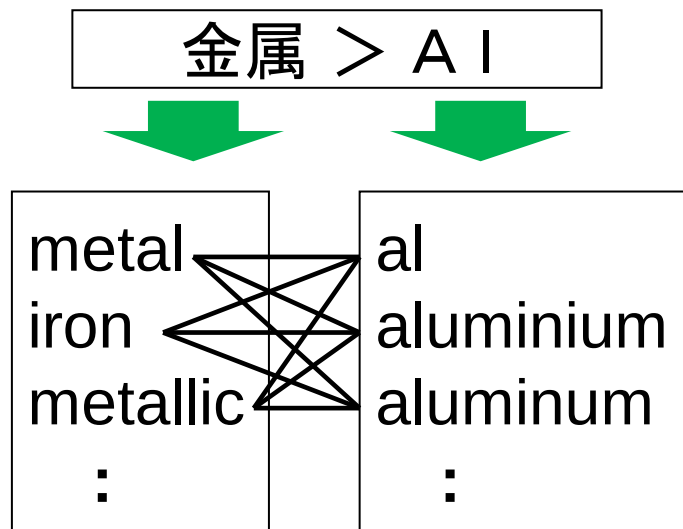
1. Translating Japanese Terms (Giza)



Translate the
Japanese hypernym
and hyponym into
English individually

Alignment of English-Japanese Hypernym-Hyponym Pairs

2. Creating English hypernym-hyponym pairs

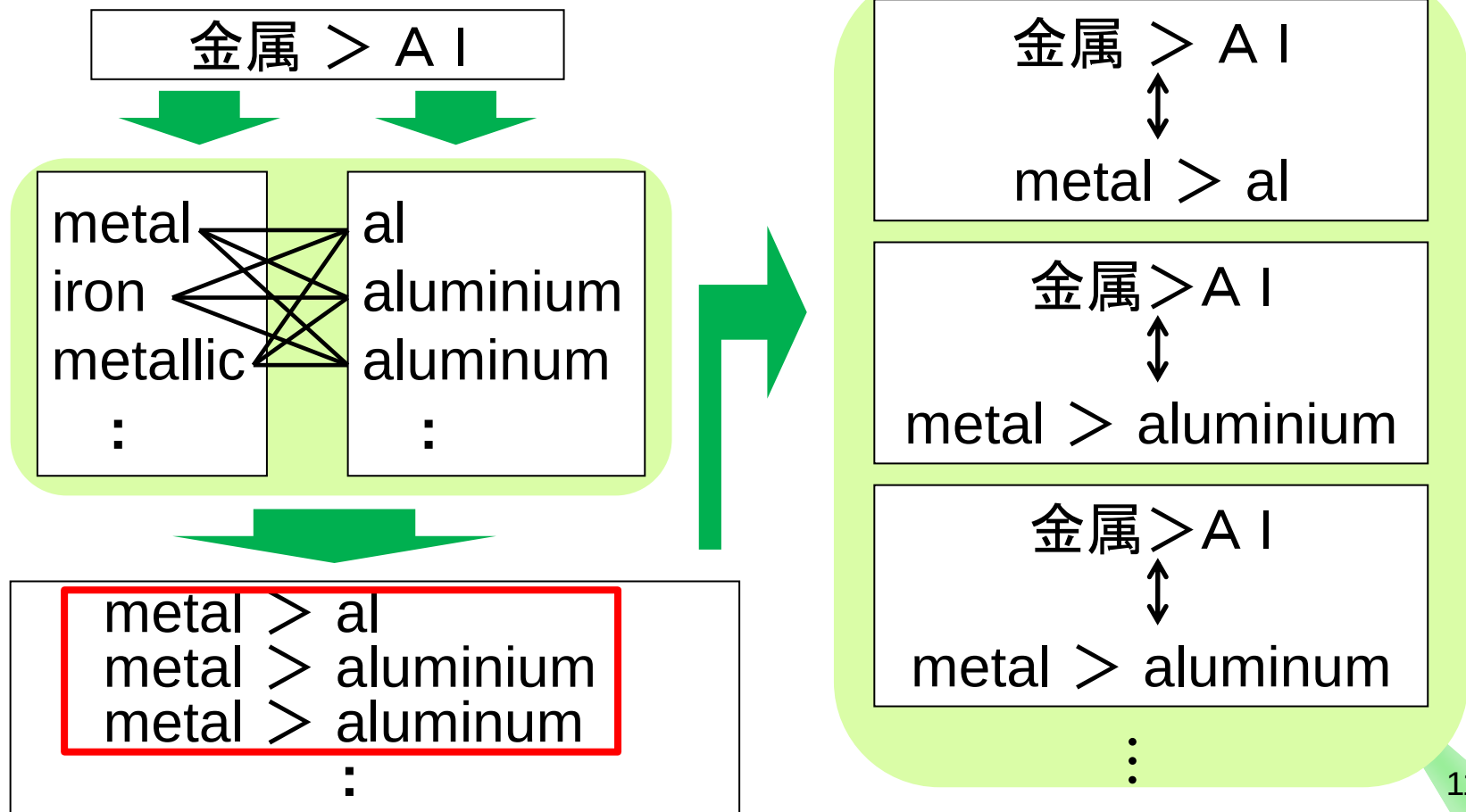


Combine them and create English
hypernym-hyponym pairs

Alignment of English-Japanese Hypernym-Hyponym Pairs

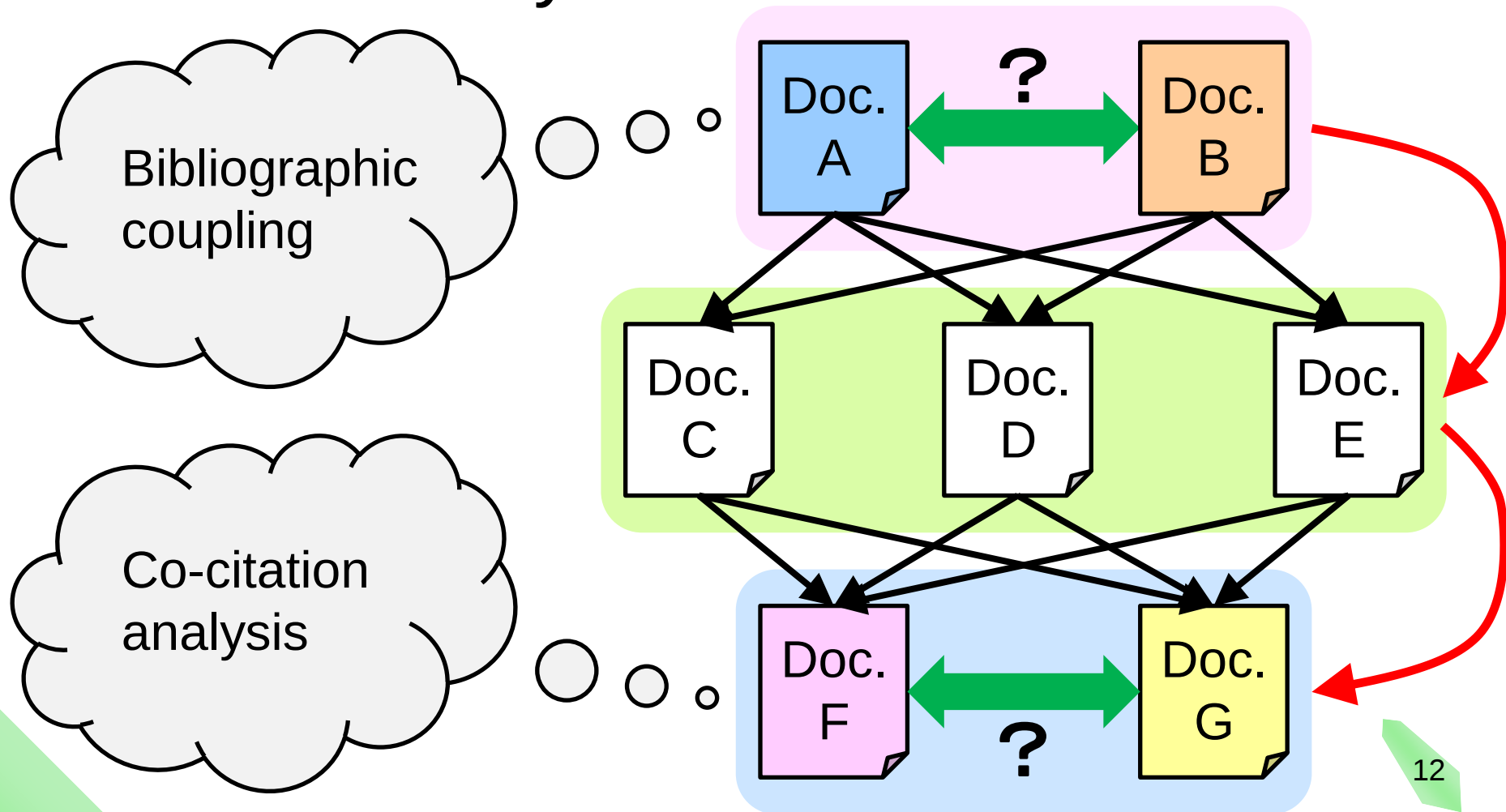
alignment

3. Aligning candidates with an English thesaurus



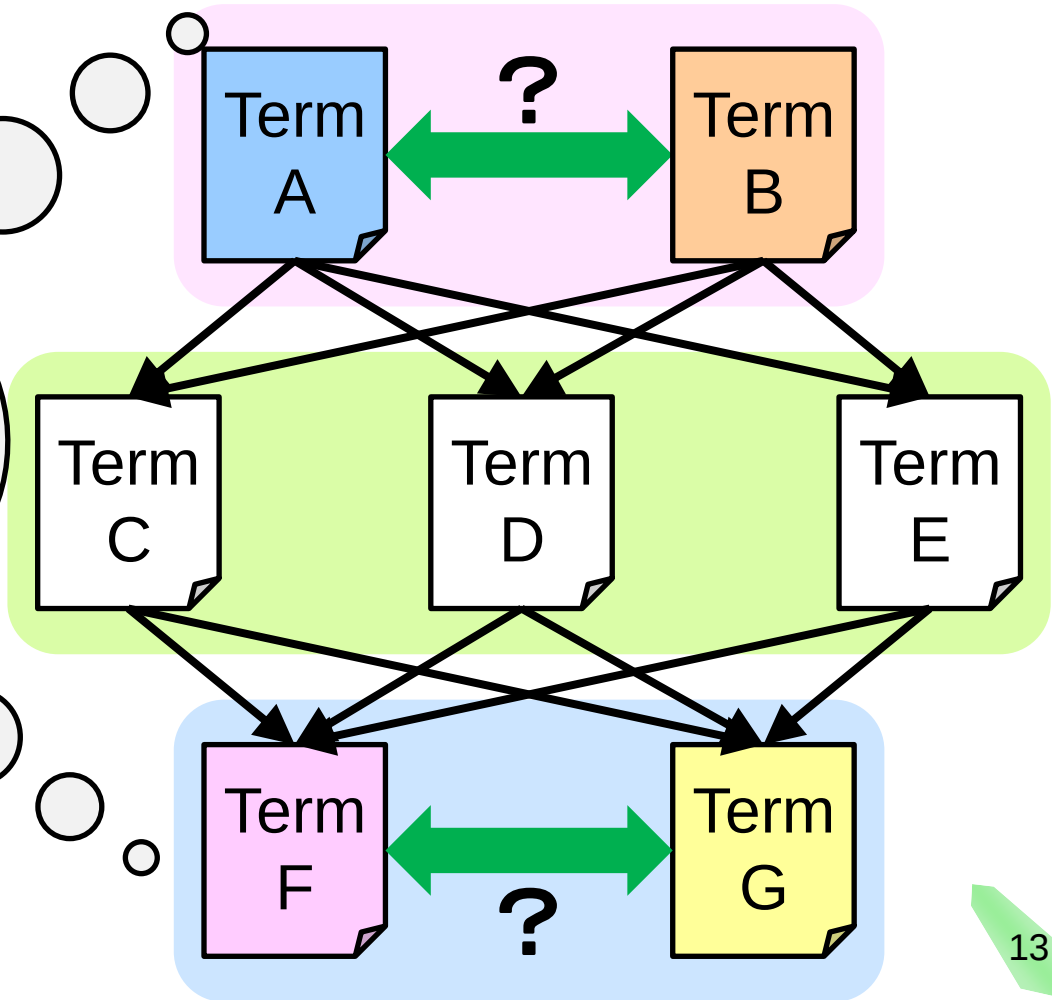
Alignment of English-Japanese Hypernym-Hyponym Pairs

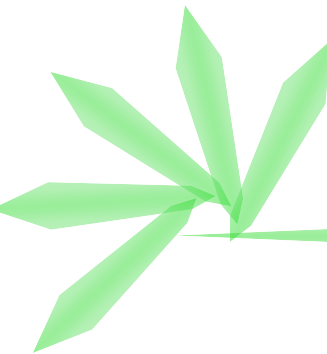
Citation analysis



Alignment of English-Japanese Hypernym-Hyponym Pairs

Consider a
hypernym-hyponym
relation as a citation
relation



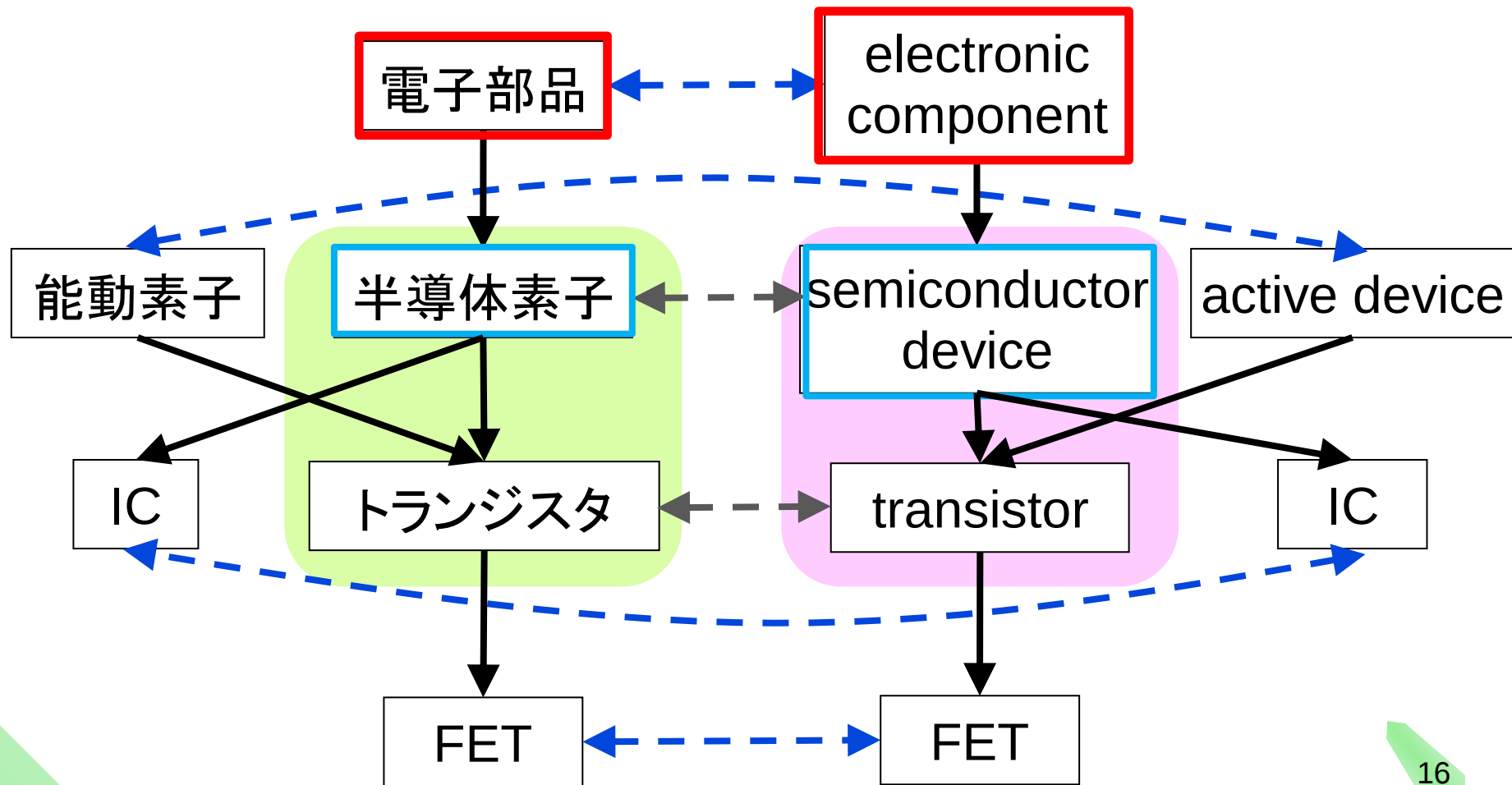


Alignment of English-Japanese Hypernym-Hyponym Pairs

- ① Translation probability
- ② Hypernyms' hypernyms
- ③ Hypernyms' hyponyms
- ④ Hyponyms' hypernyms
- ⑤ Hyponyms' hyponyms

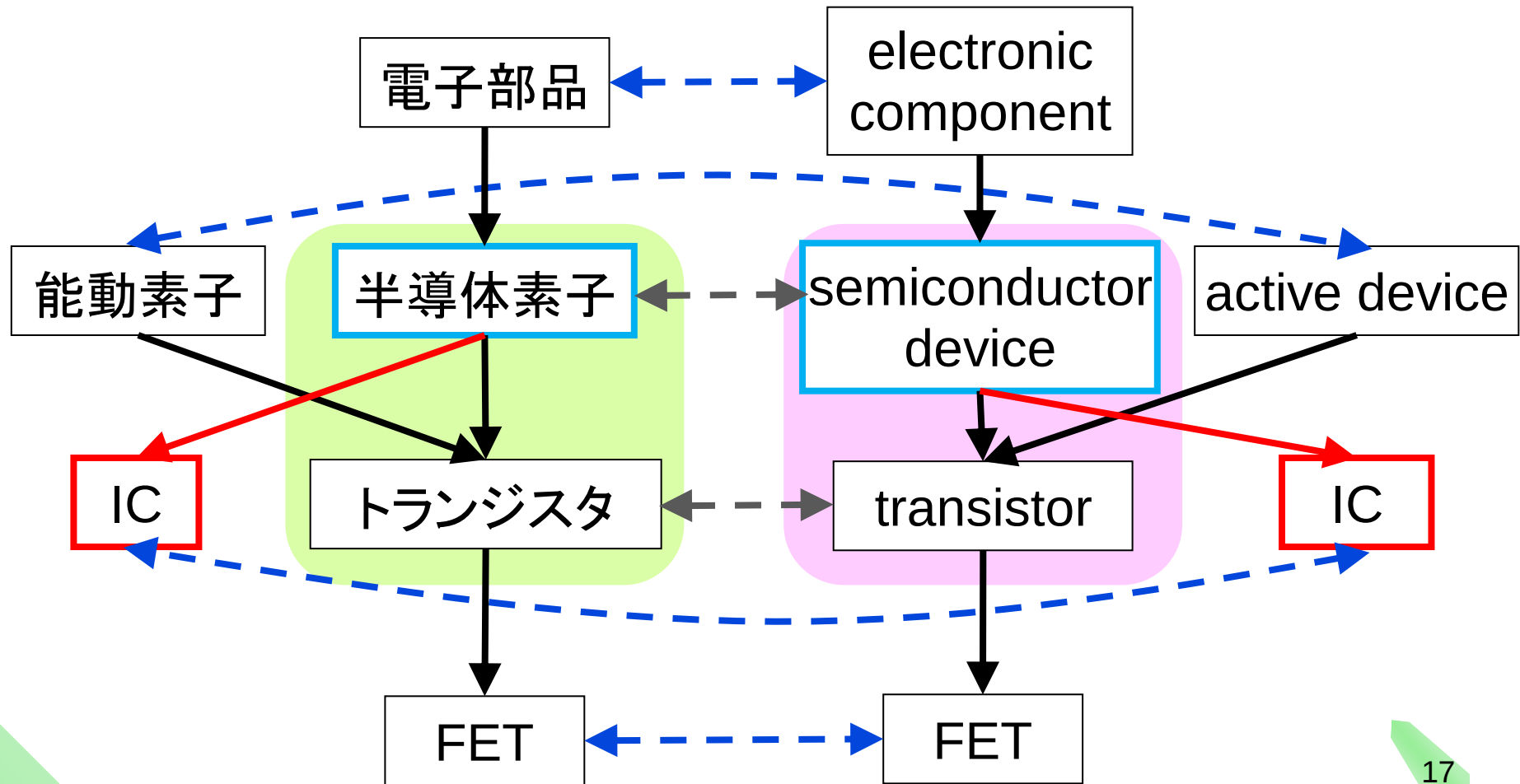
Alignment of English-Japanese Hypernym-Hyponym Pairs

② Hypernyms' hypernyms



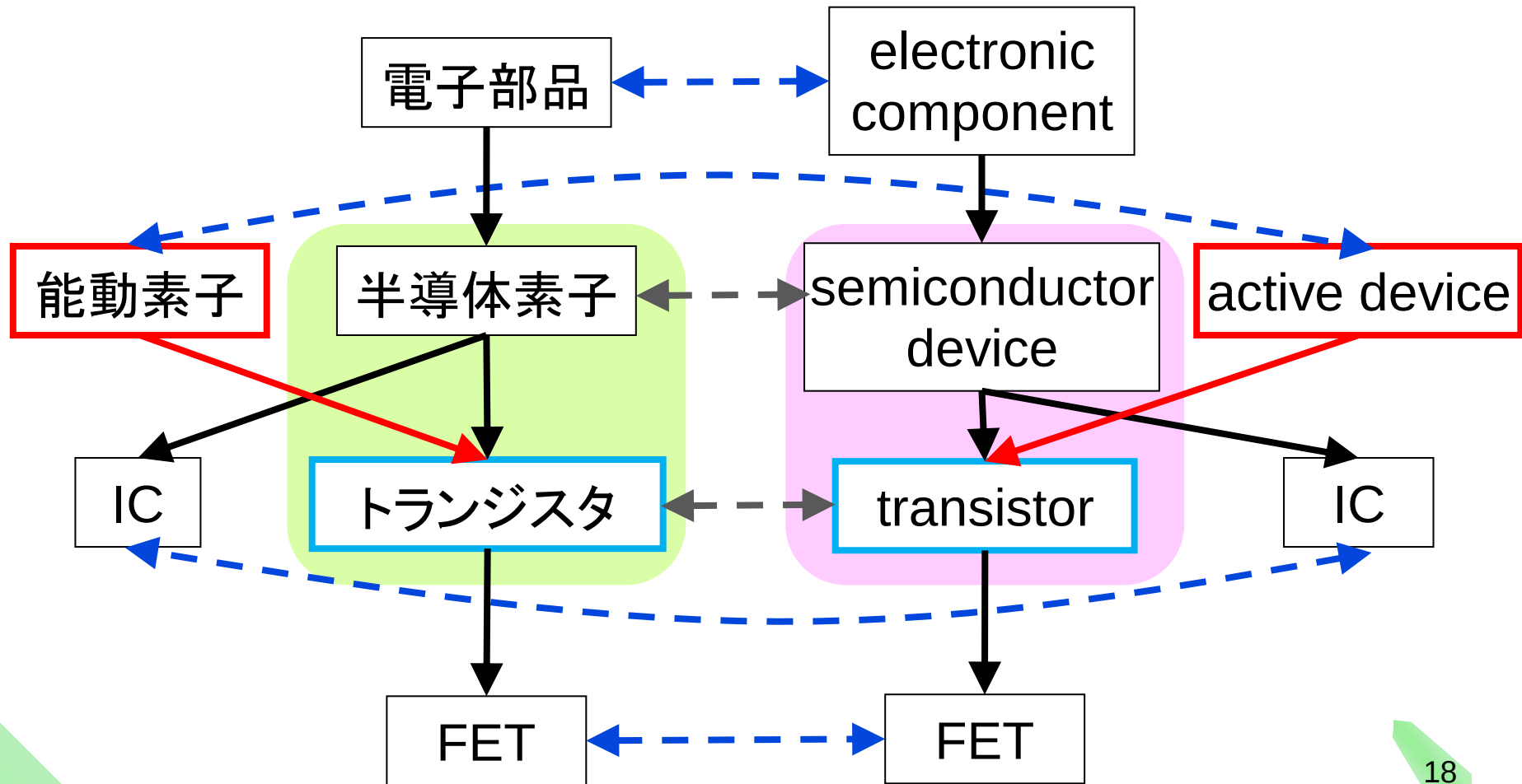
Alignment of English-Japanese Hypernym-Hyponym Pairs

③ Hypernyms' hyponyms



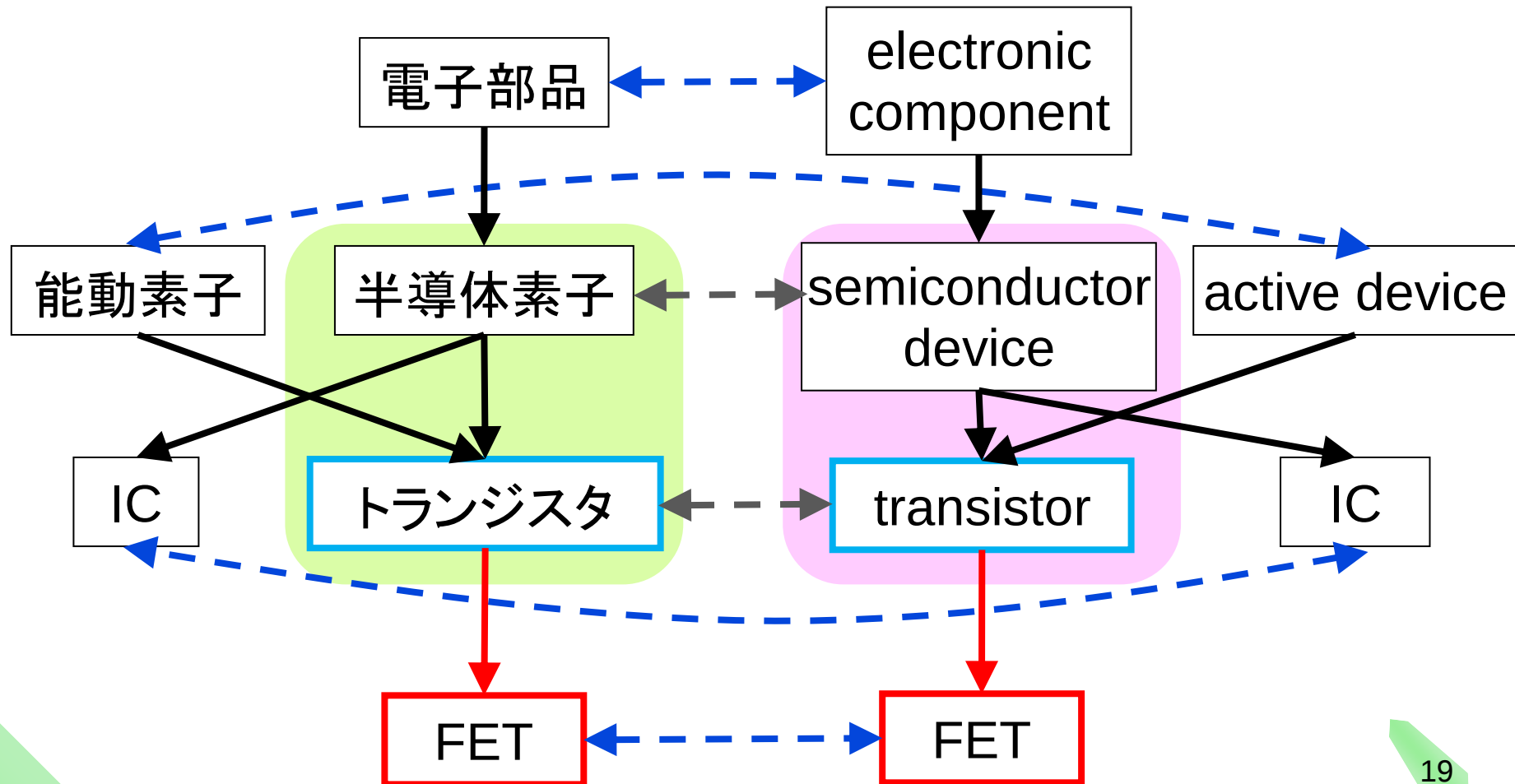
Alignment of English-Japanese Hypernym-Hyponym Pairs

④ Hyponyms' hypernyms



Alignment of English-Japanese Hypernym-Hyponym Pairs

⑤ Hyponyms' hyponyms





Alignment of English-Japanese Hypernym-Hyponym Pairs

① Translation probability

Hypernym

$$P(J \Rightarrow E) \times P(E \Rightarrow J)$$

×

Hyponym

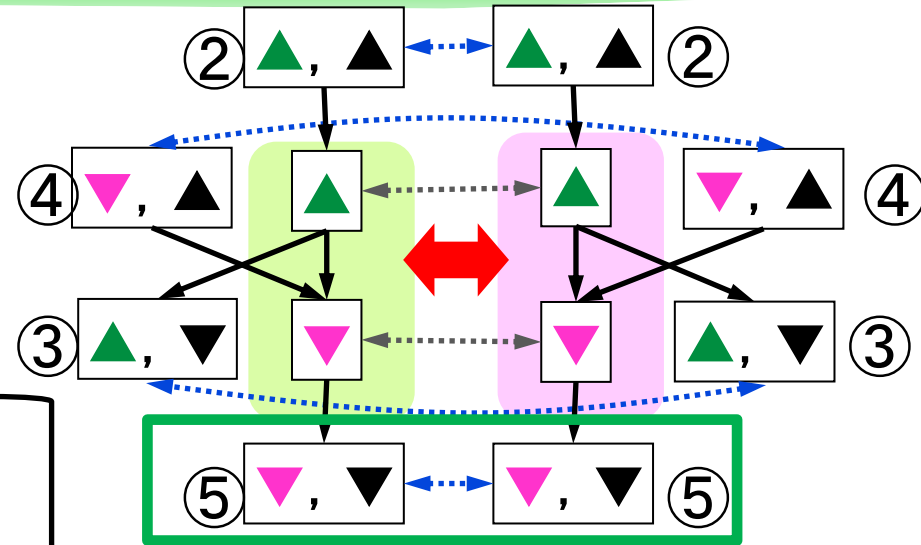
$$P(J \Rightarrow E) \times P(E \Rightarrow J)$$

Alignment of English-Japanese Hypernym-Hyponym Pairs

Features

- ① Translation probability
- ② Hypernyms' hypernyms
- ③ Hypernyms' hyponyms
- ④ Hyponyms' hypernyms
- ⑤ Hyponyms' hyponyms

Select a feature



Combine ① and
②~④

four methods



Evaluation Procedure

1. Calculation of the validity of each candidate

a: value of feature ①

b: value of feature ②~⑤

$$a^{\beta} + \alpha \times b \quad \left[\begin{array}{l} \beta: 1/5, 1/10, 1/15, 1/20 \\ \alpha: 0.1, 0.2, \dots, 0.9 \end{array} \right]$$

2. Parameter tuning

x: threshold value

$$\left[x: 0.00, 0.01, 0.02, \dots, 2.00 \right]$$

3. Evaluation

Evaluate our methods with the threshold value using a test data set



Experiments

Data

- 3,898,060 relations from USPTO patents
- 7,031,159 relations from Japanese patents

manually evaluated pairs of English-Japanese
hypernym-hyponym relations

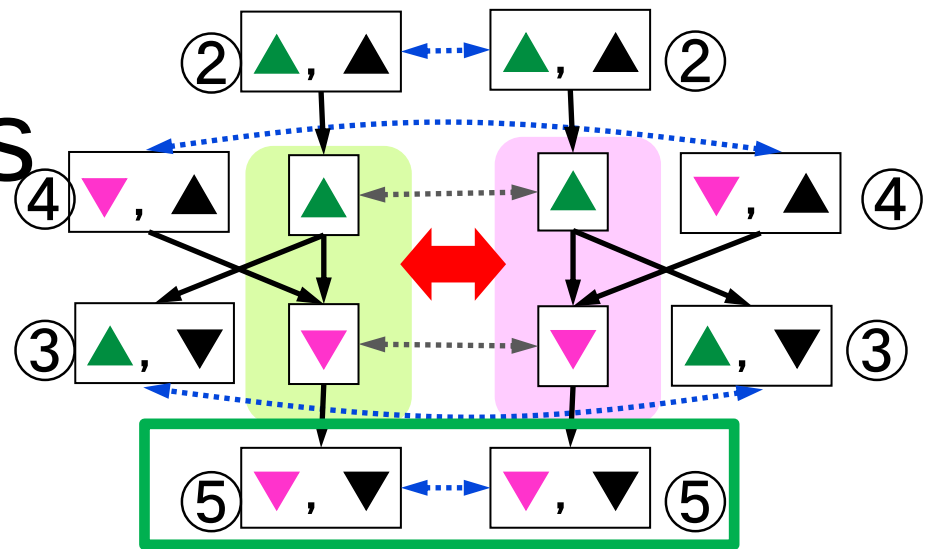
Correct	Incorrect	Total
982	1,653	2,635

Evaluation

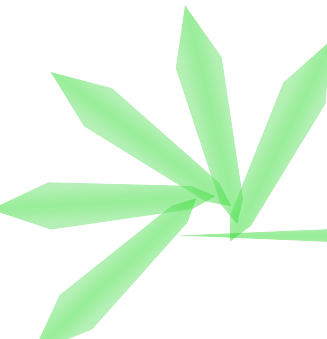
Recall, Precision, and F-measure

Experiments

Alternatives



		①	②	③	④	⑤
Baseline		○				
Our method	a	○	○			
	b	○		○		
	c	○			○	
	d	○				○



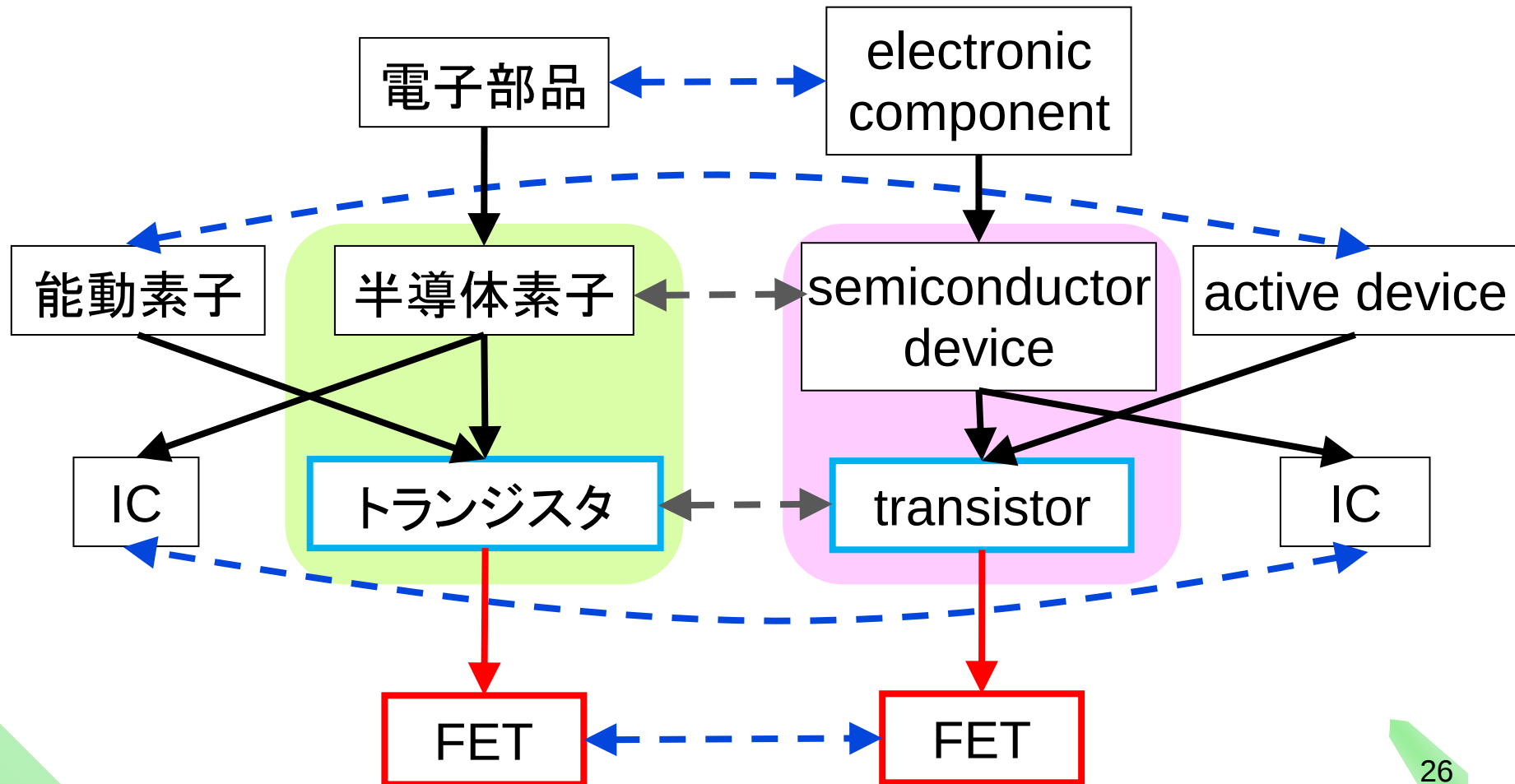
Experimental Results

		α	β	Precision	Recall	F-measure
Baseline (e)		0	1/15	78.5	77.8	78.0
Our Meth ods	a	0.1	1/10	76.4	78.1	77.1
	b	0.1	1/20	76.3	79.5	77.4
	c	0.1	1/15	75.8	78.4	76.9
	d	0.1	1/15	77.5	79.4	78.3

Our method (d) obtained the best score

Alignment of English-Japanese Hypernym-Hyponym Pairs

⑤ Hyponyms' hyponyms



Discussion (1/3)

Improvement of Recall Values

記録媒体 > CD



record medium > cd

車両 > 自動車



vehicles > car

A baseline method could align these hypernym-hyponym pairs, because translation probabilities are low.



Our method could align them!



Discussion (2/3)

Cases that our method mistakenly aligned

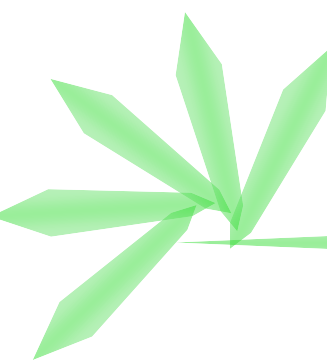
Japanese		English	
Hypernym	Hyponym	Hypernym	Hyponym
金属 (metal)	亜鉛 (zinc)	metal	aluminum- zinc
着色剤 (coloring agent)	染料 (dye)	coloring agent	organic dye



Discussion (3/3)

Cases that our method could not align

Japanese		English	
Hypernym	Hyponym	Hypernym	Hyponym
金属材料 (metallic material)	銅 (copper)	metallic material	Cu
電子部品 (electric part)	I C (IC)	electric part	integrated circuit



Conclusion

Propose a method for constructing a bilingual thesaurus using citation analysis.

(d) Hyponym's hyponyms

Recall: 77.5% Precision: 79.4%
F-measure: 78.3 %