公開用抜粋版

筑波大学
University of Tsukuba

文書データからの知識発見

2021/02/05

筑波大学大学院ビジネスサイエンス系
津田和彦

公開用抜粋版

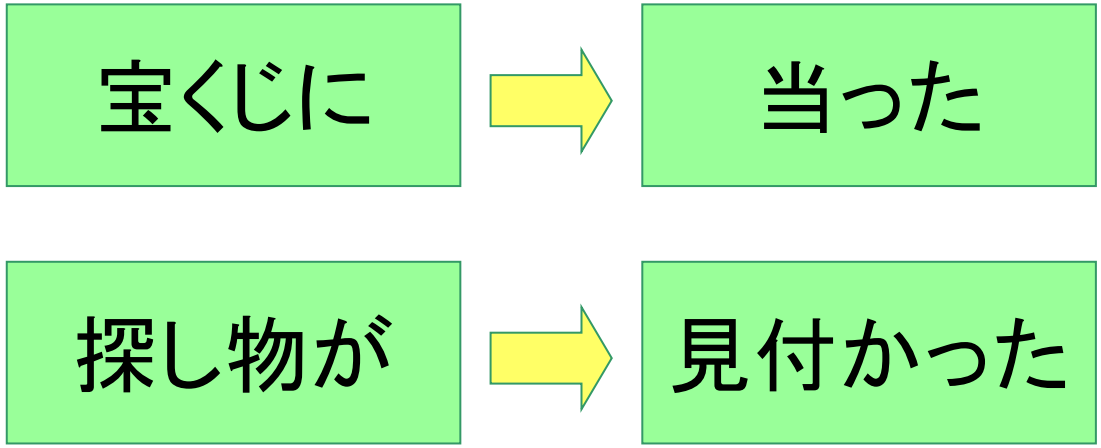
筑波大学
University of Tsukuba

テキストマイニング技術

動詞の曖昧性

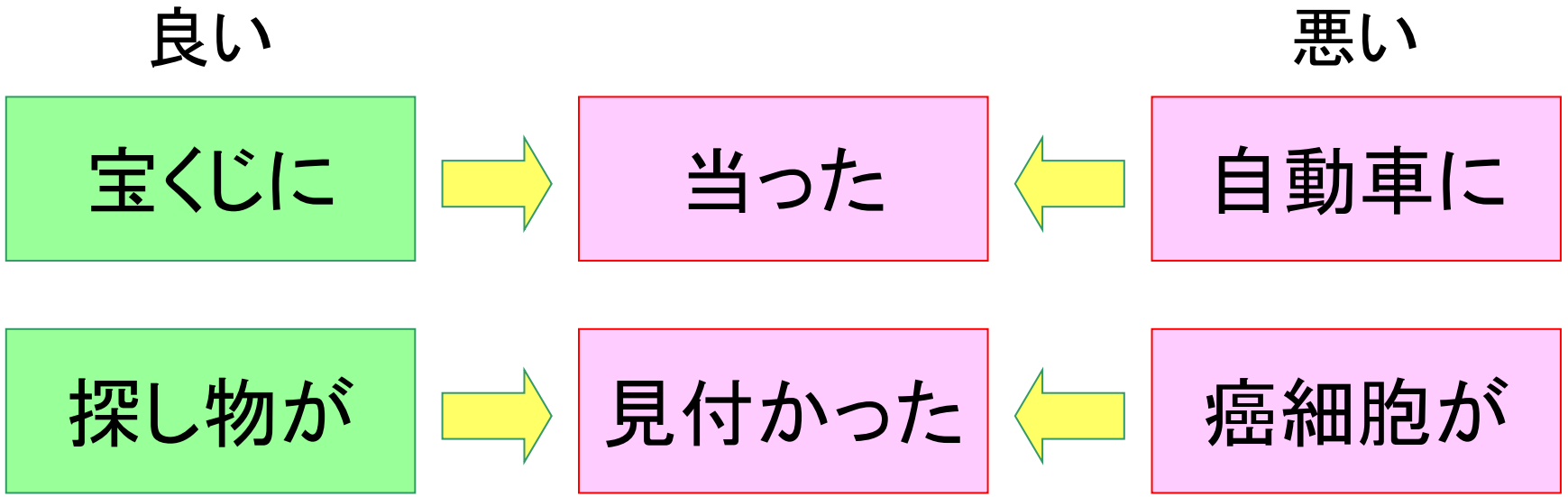
- 文の基本は，主語（名詞）＋述語（動詞）

良い



動詞の曖昧性

- 文の基本は，主語（名詞）＋述語（動詞）



テキストマイニング

1. 私は，学校の先生です．

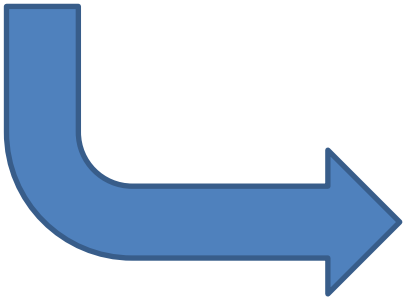
   Keyword=（私，学校，先生）

2. 私は，学校の先生の父が居ます．

   Keyword=（私，学校，先生，父）

3. 私は，大学の教官です．

   Keyword=（私，大学，教官）



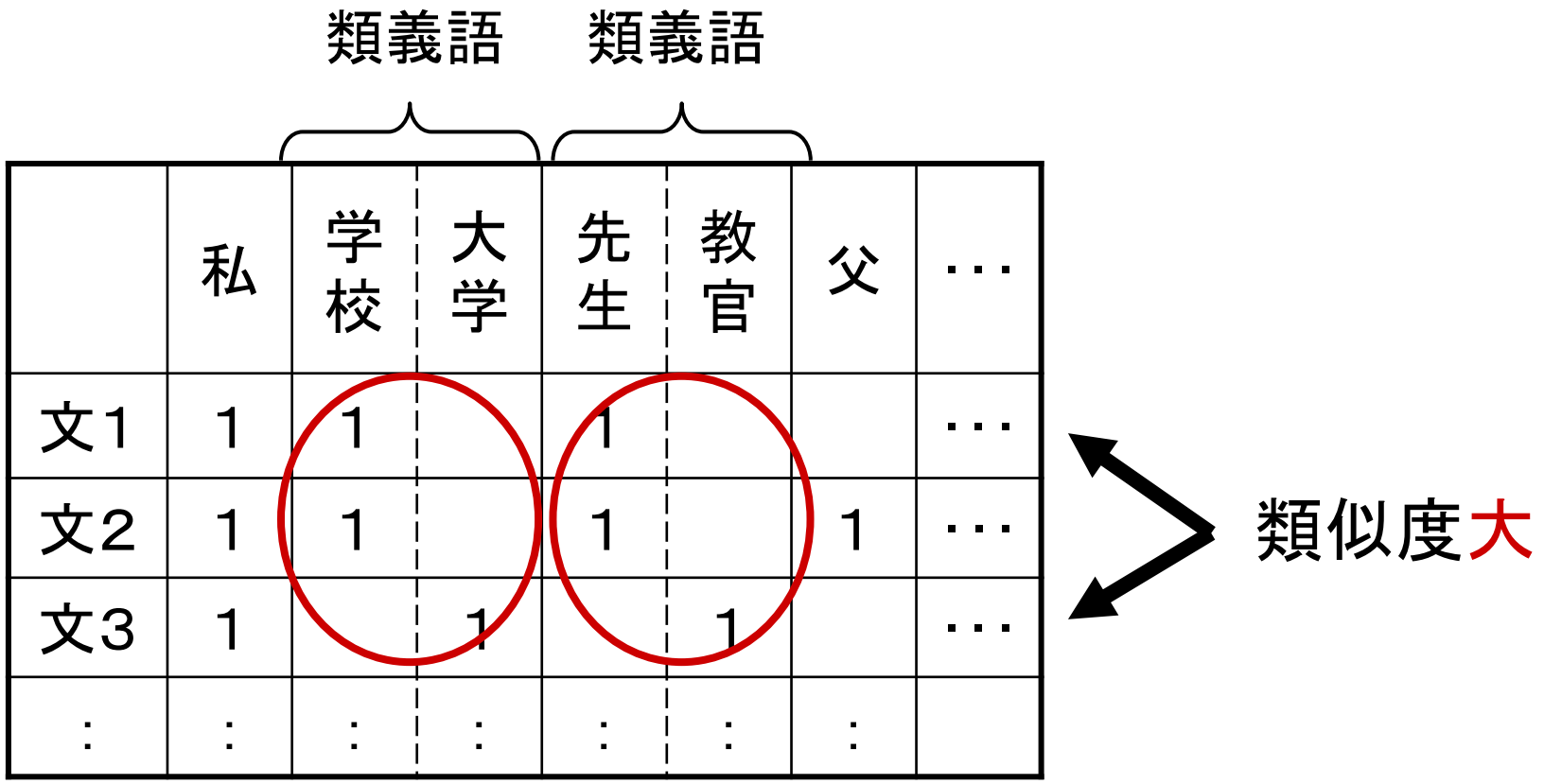
	私	学校	先生	父	大学	教官	…
文1	1	1	1				…
文2	1	1	1	1			…
文3	1				1	1	…
:	:	:	:	:	:	:	

テキストマイニング

	私	学校	先生	父	大学	教官	…
文1	1	1	1				…
文2	1	1	1	1			…
文3	1				1	1	…
:	:	:	:	:	:	:	

類似度大

テキストマイニング



公開用抜粋版

筑波大学
University of Tsukuba

テキストマイニングの評価

評価指標

- Recall　再現率

$$Recall = \frac{\text{検出した正解数}}{\text{全正解数}}$$

- Precision　適合率／精度

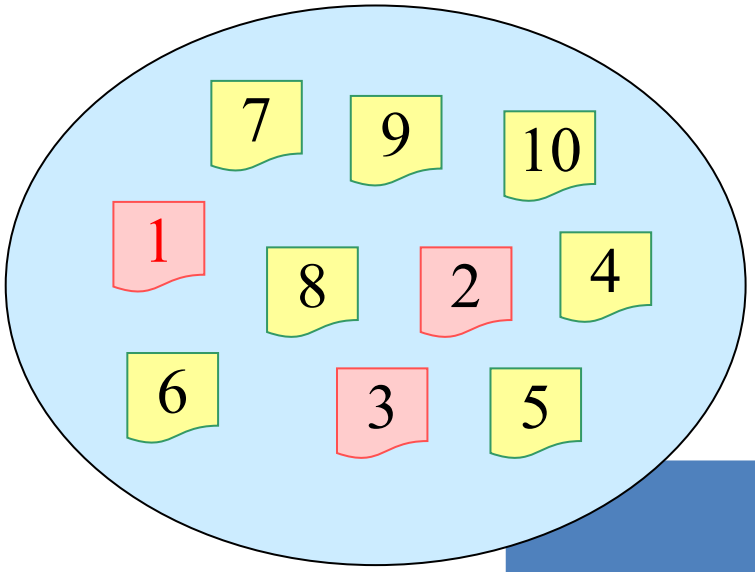
$$\text{Precision} = \frac{\text{正解数}}{\text{検出した数}}$$

- F-measure F-尺度／F値

$$F - measure = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

- 新奇性率

評価結果



1～3が正解で，7～10が誤り

	Recall	Precision	F-measure
1 2 3	1.00	1.00	1.00
1 2 3 5 7	1.00	0.60	0.75
1 3 7 8	0.67	0.50	0.57
2 3 5	0.67	0.67	0.67

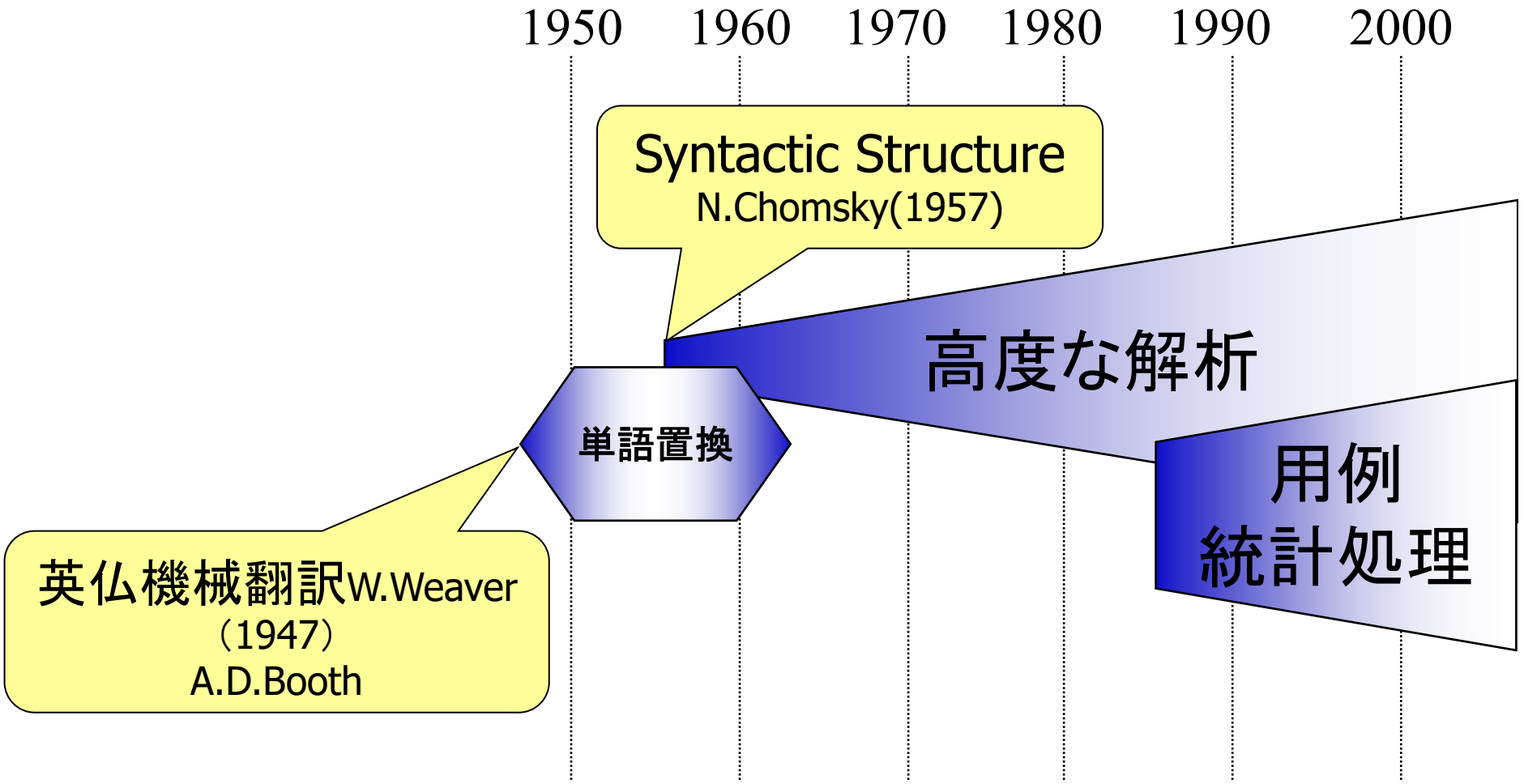
公開用抜粋版

筑波大学
University of Tsukuba

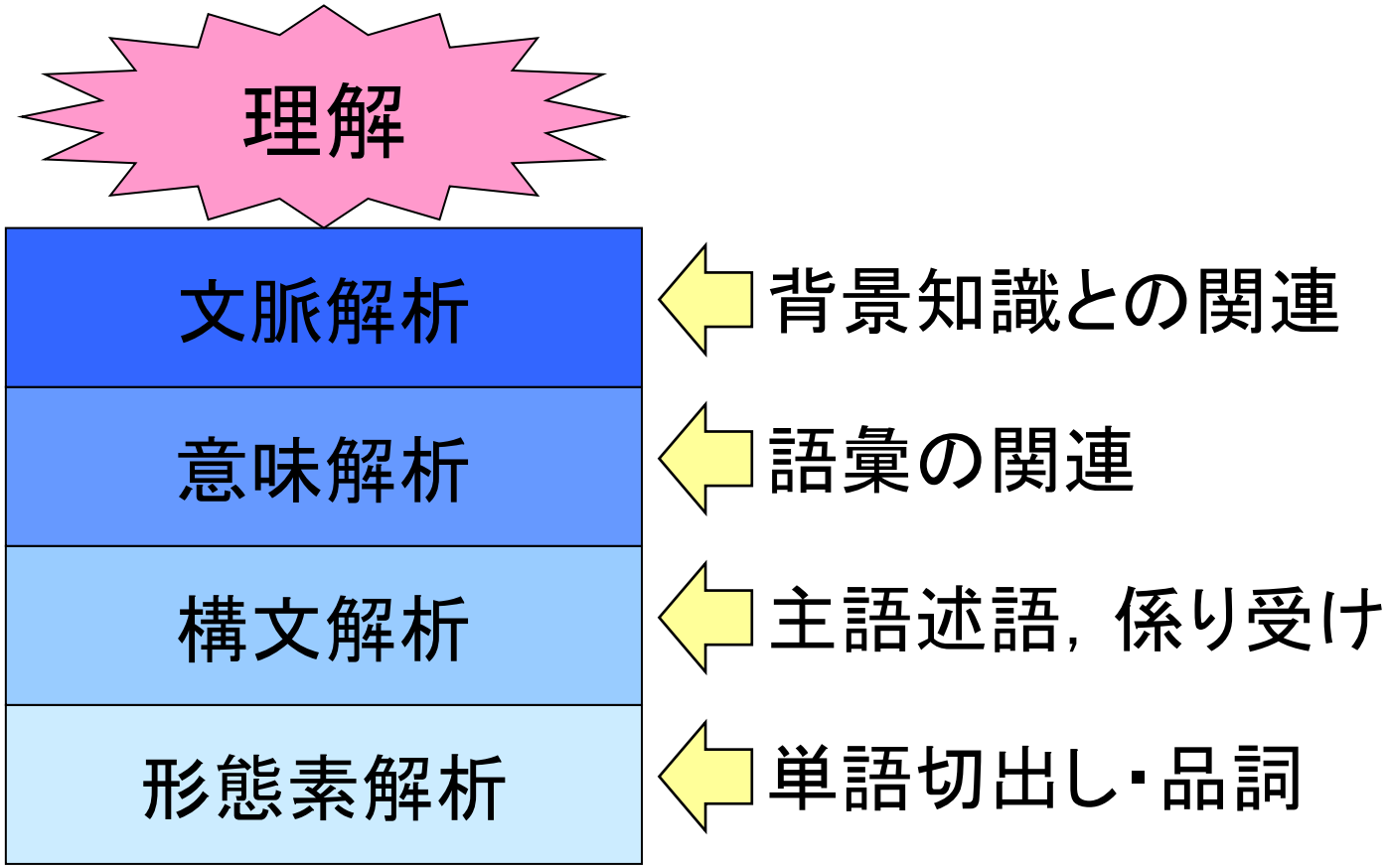
自然言語処理技術の歴史と基礎

言語処理の歴史

＝機械翻訳の歴史



日本語の解析手法



格文法

- 動詞を中心に，助詞により解析