

係り受け解析を考慮した企業のIR情報からの戦略フレーズ抽出方法の提案

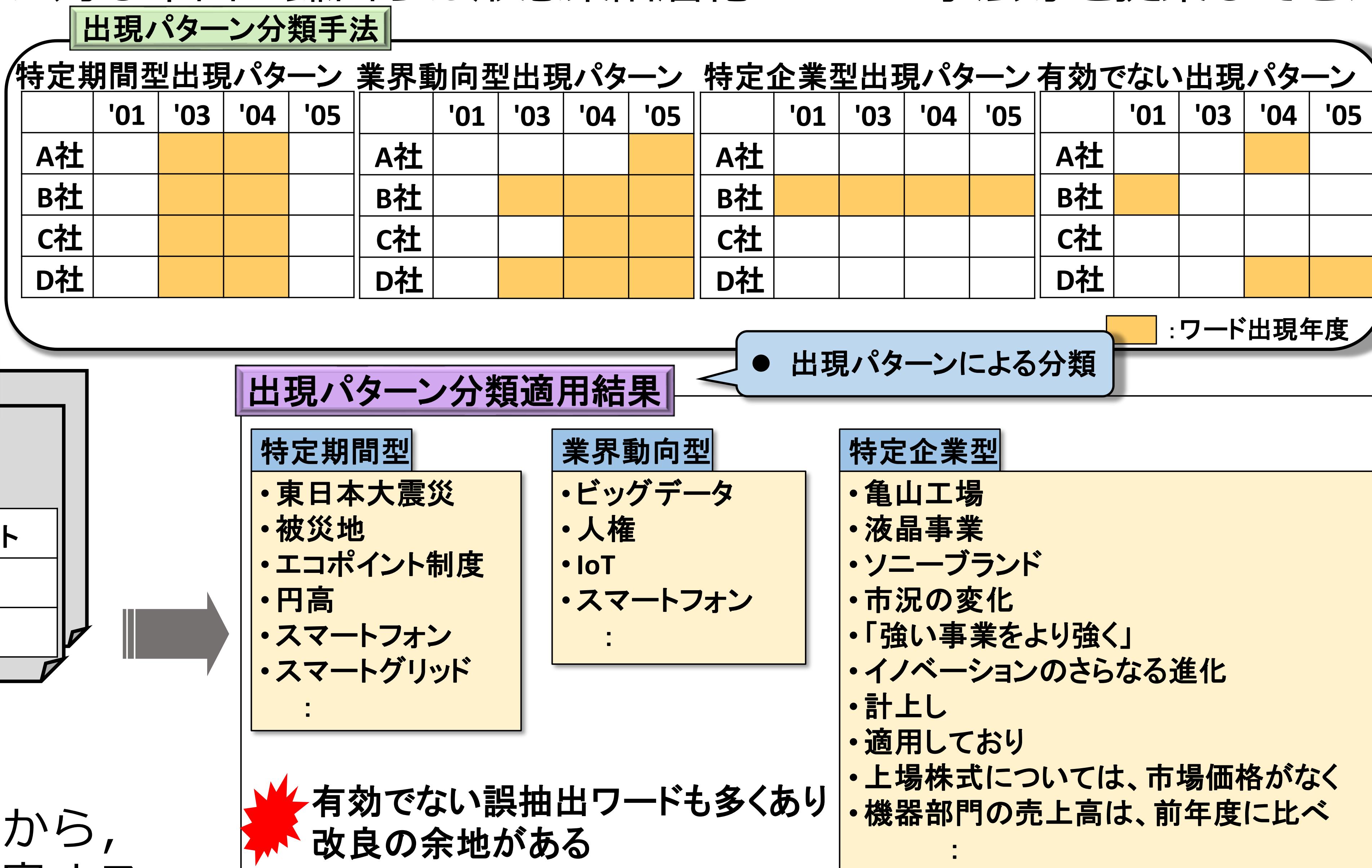
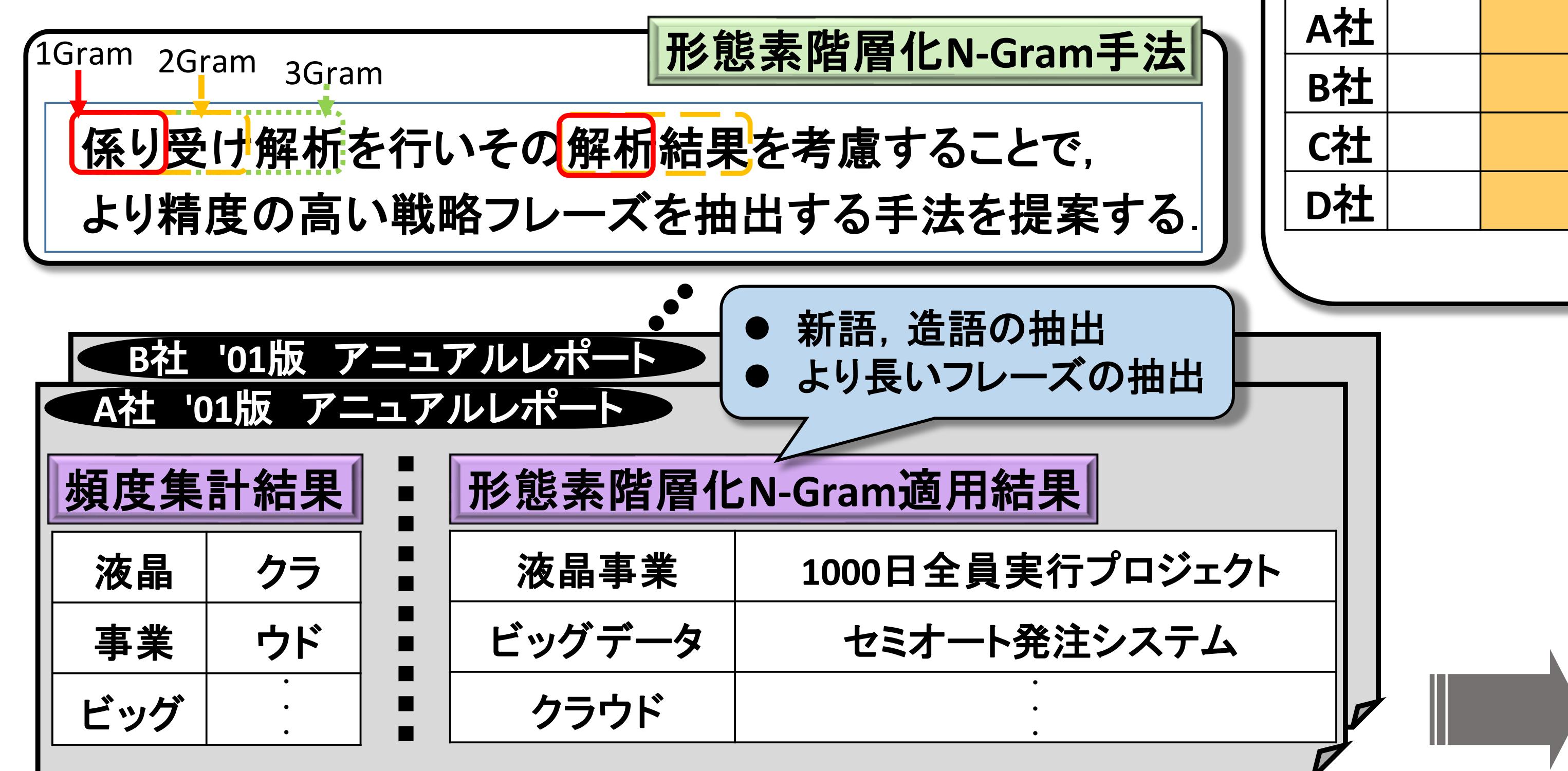
A Proposal of a Method for Extracting Strategy Phrases Considering with Dependency Analysis from IR Documents of Enterprises

日出間 健(東京都市大学 大学院) 岡田 公治(東京都市大学)

1. 背景

企業は多くのIR情報を開示している。IR情報には戦略や経営施策等が記述され、他企業と比較することで有益な情報を得ることができる。IR情報に対してテキストマイニングを適用する研究もあるが一般的な形態素の頻度集計等では戦略や経営施策を表す新語(例えば「クラウド」「ビッグデータ」等)やフレーズ(「イノベーションのさらなる進化」等)の抽出は難しい。これに対し峯田・鎌田らは形態素階層化N-Gram手法等を提案してきた。

2. 従来手法(峯田・鎌田ら)



3. 目的

従来手法を用いて抽出された戦略ワードから、誤抽出ワードを取り除く手法を新たに提案する。

4. 従来手法による抽出結果

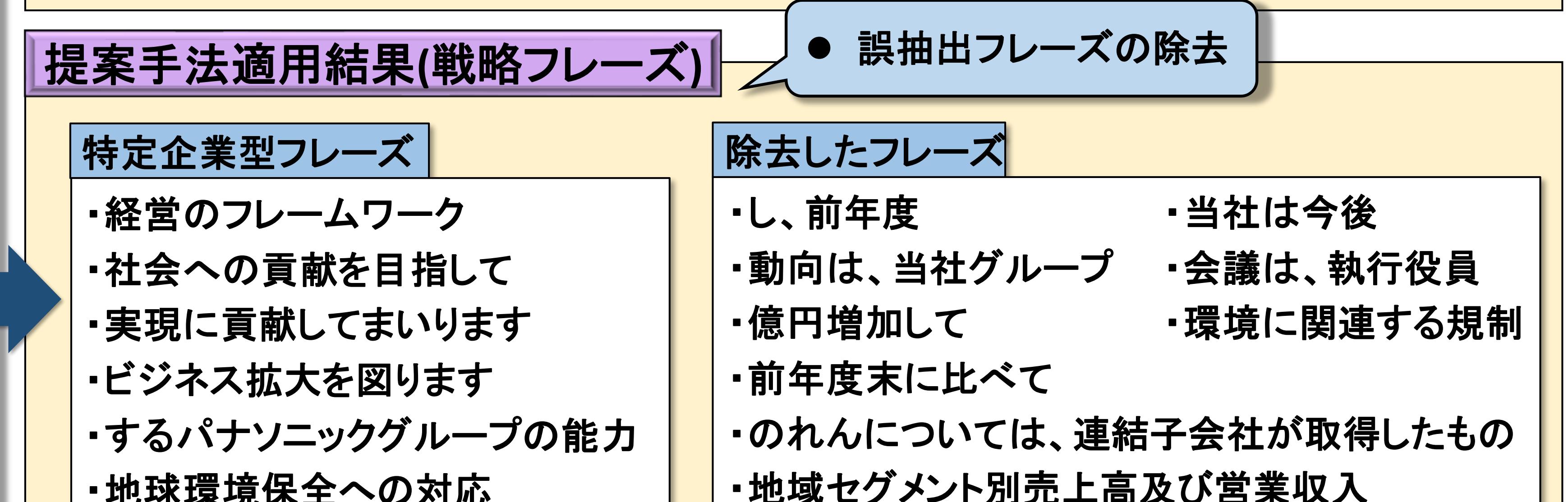
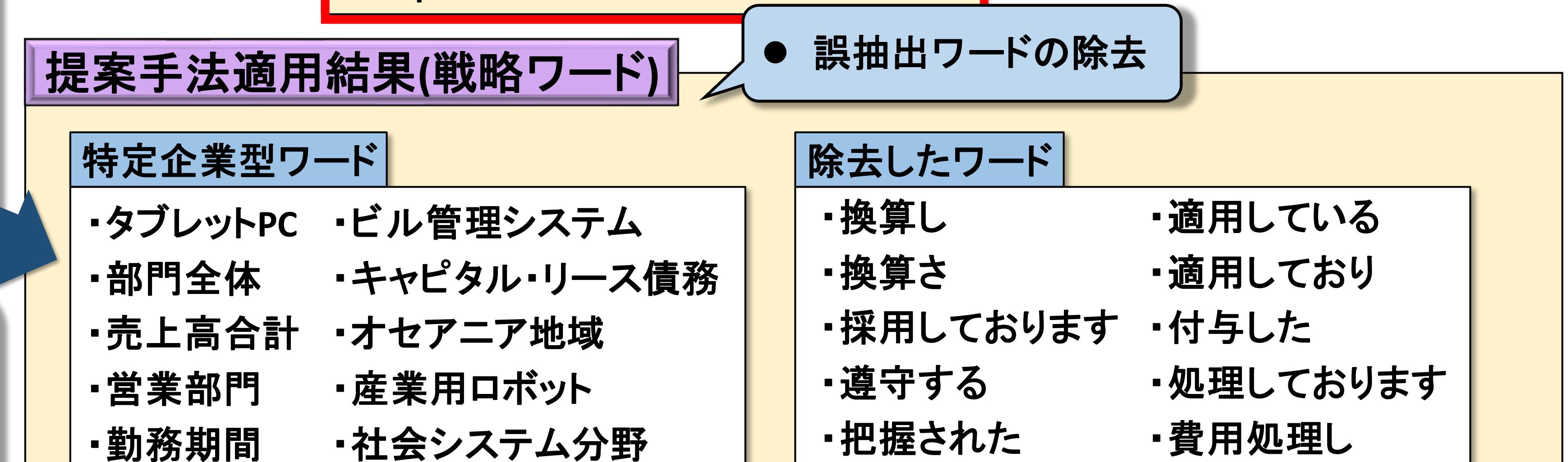
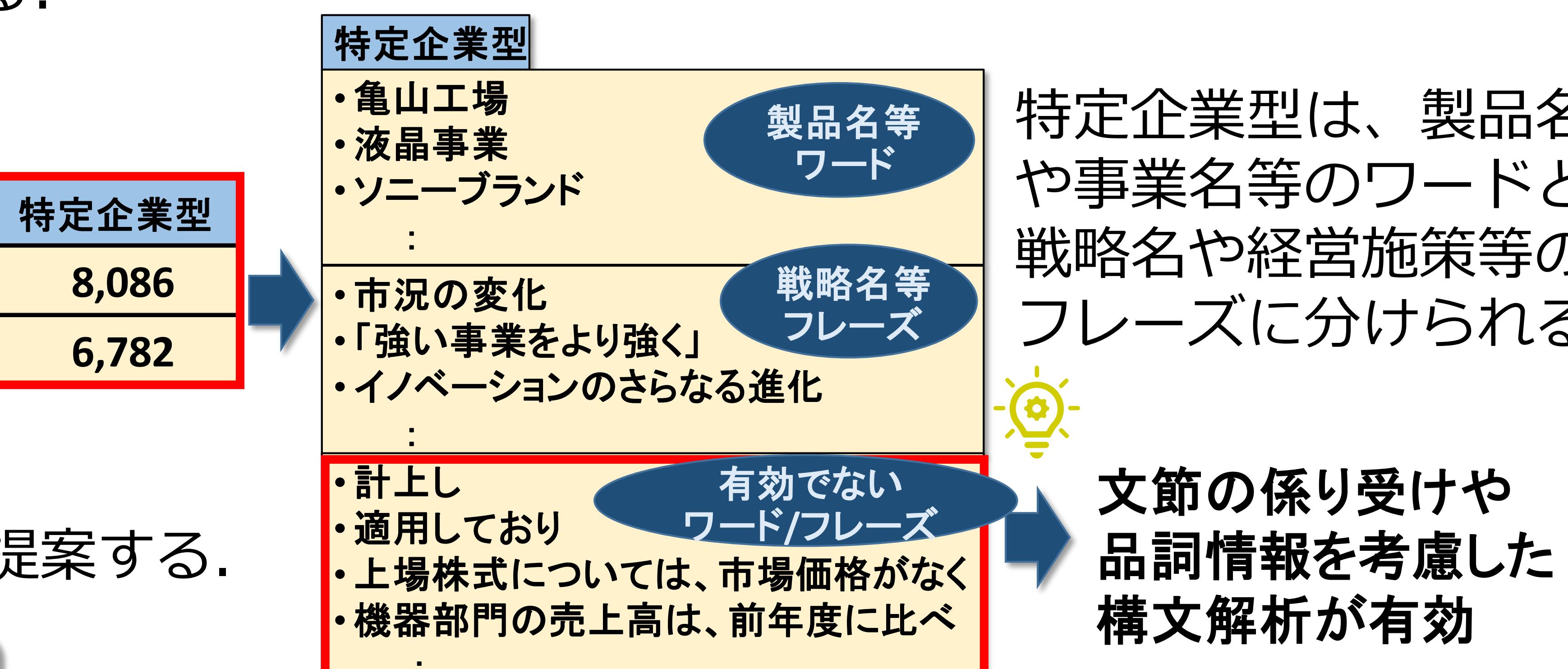
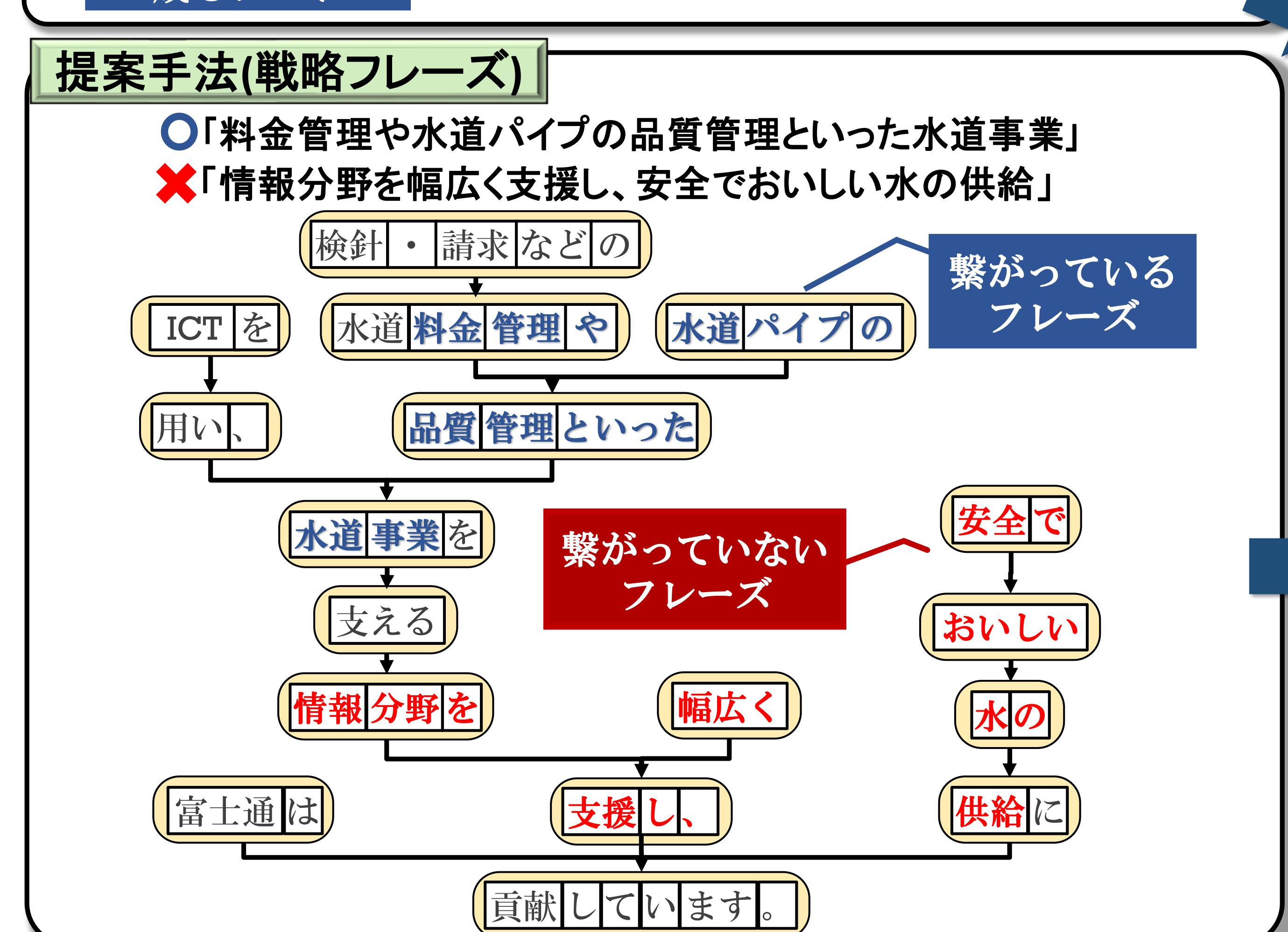
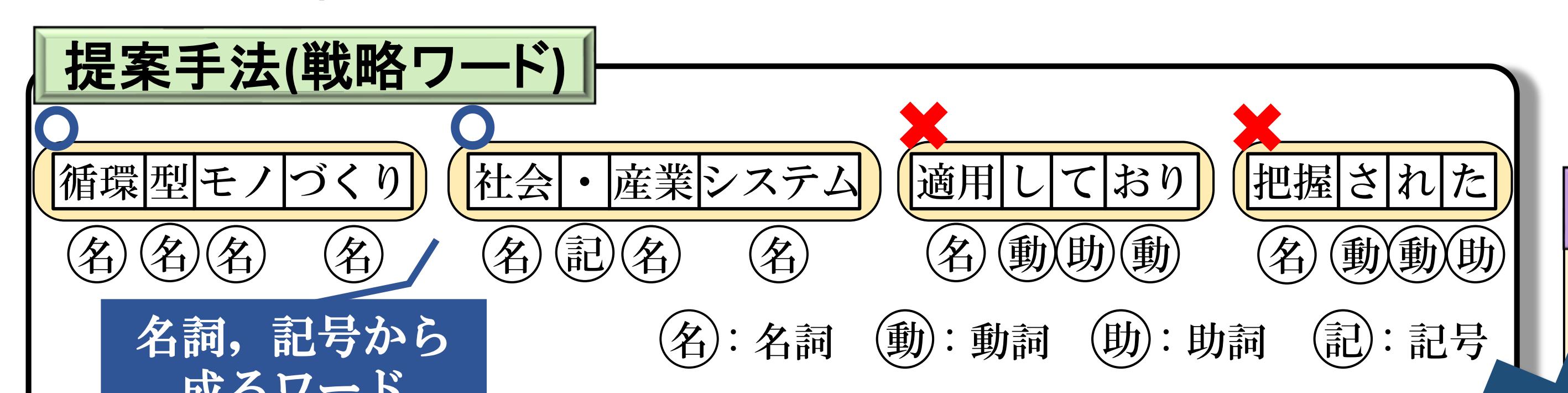
特定企業型の数が多く、誤抽出ワードを多く含んでいる。

	特定期間型	業界動向型	特定企業型
鉄道業界	5	0	8,086
電機業界	15	6	6,782

特定企業型は、製品名や事業名等のワードと戦略名や経営施策等のフレーズに分けられる。

5. 新たな抽出手法の提案

係り受け構造や品詞情報を考慮した抽出手法を提案する。



	鉄道業界		電機業界	
従来手法	8,086		6,783	
提案手法	戦略ワード	戦略フレーズ	戦略ワード	戦略フレーズ
	4,151	2,710	3,117	2,486

従来手法の抽出結果からは、特定企業型に誤抽出ワードが多いこと、ワードとフレーズに分類でき抽出したい情報が異なることが分かった。そこで、構文解析によって品詞情報や係り受け構造から誤抽出を省く手法を提案した。また、提案手法を複数業界に適用し有効性の確認を行った。

6. 結論