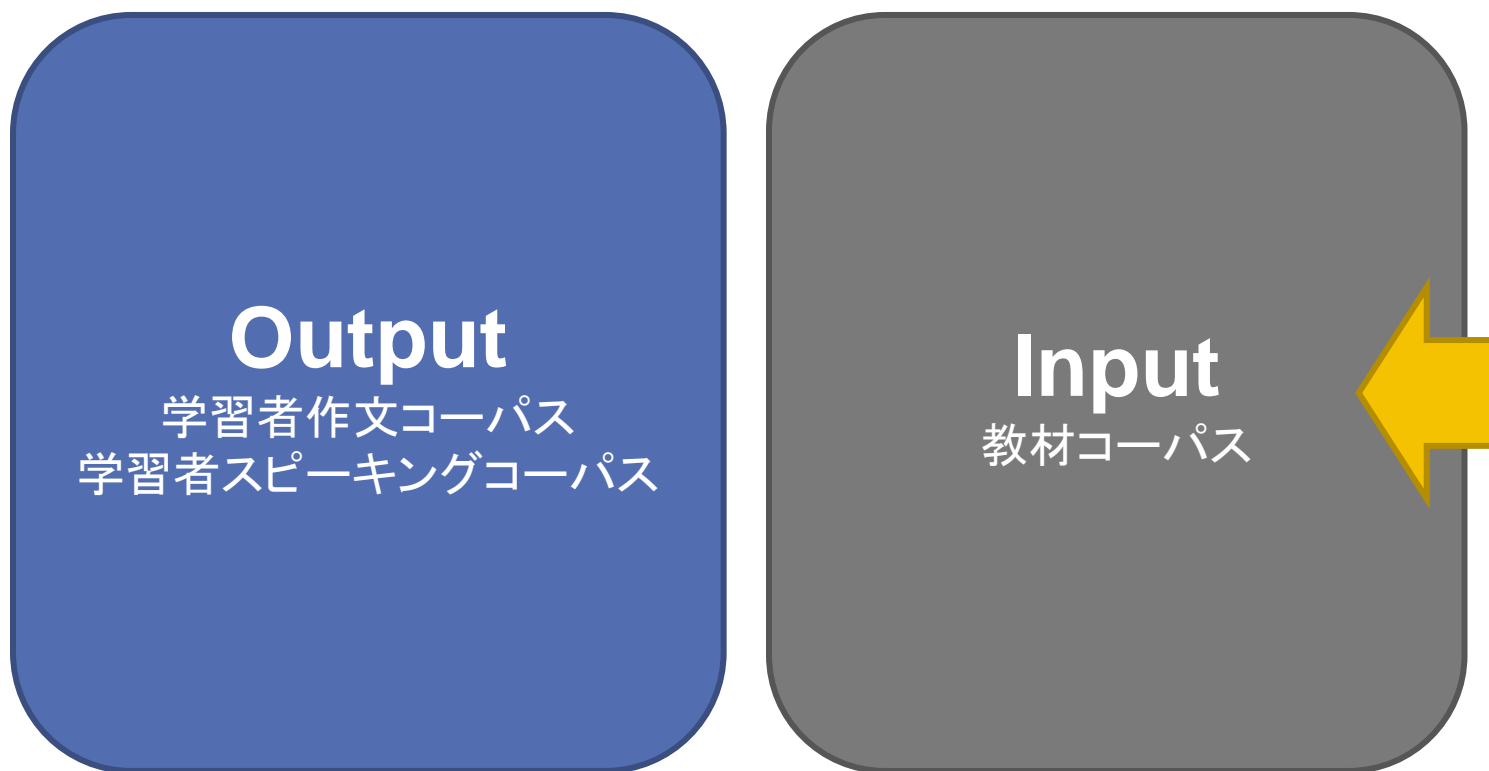


# 教材コーパスの構築経緯： テキスト選定に関する中間報告

2013/03/21

# はじめに: 教材コーパスの位置づけ

## 学習者の「言語」



レベル別に「言語特徴」を抽出する

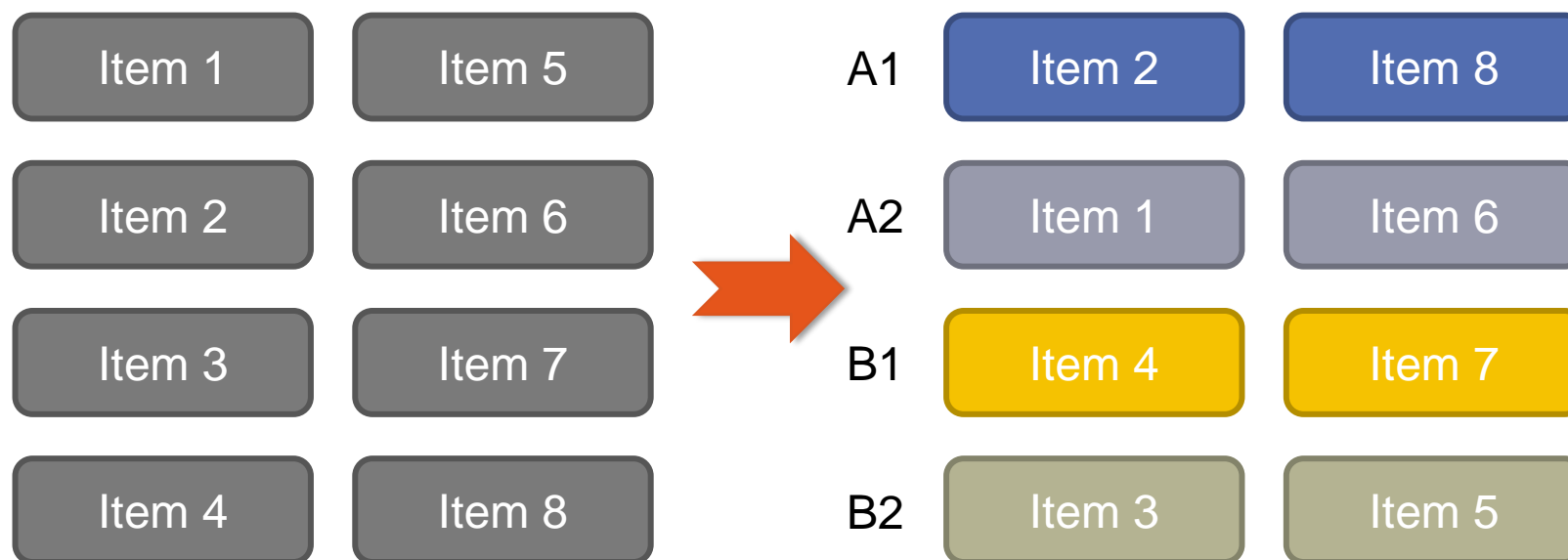
# 教材コーパス分析の目的

## 【目的】

**CEFRLレベルに基づいた教材**を用いてレベル決定に作用している「**言語的特徴**」を抽出する

# 教材コーパス分析の目的

【イメージ1】 言語項目をレベルごとに貼り付ける



【イメージ2】 合成変数で説明する

$$\text{CEFRレベル} = \text{V-score} + \text{G-score} + \text{C-score} \dots$$

# 投野科研教材コーパス班の特徴

## Data-drivenなアプローチ

→ **CEFRベースで編纂されたテキスト** (レベル付け済み) を中心に量的に十分な程度コーパス化し、言語特徴を抽出する

### ・inputデータを用いる

→ cf. 学習者コーパス班、English Profile

### ・CEFR-Jへの対応を探る (Pre-A1, A1 × 3レベル)

→ 日本人の80%以上がAレベル (根岸他2010)

# コーパス化の対象となる教材

- ・CEFRに基づいて作られたものを対象とする

Appropriate for the Common European Framework Level C1



- 既存のものをCEFRレベルに当てはめたものは除外
- A1-A2などレベルにまたがるものは除外

# コーパスのタグ付け(計画段階)

## ・品詞タグ

→動詞、名詞、形容詞...

## ・カテゴリータグ(一部)

### 【4技能】

Reading / R

Writing / W

Speaking / S

Listening / L

【語彙】 ※構文処理、改行処理が特殊になるため、他と区別する

WritingV / WV (書くための文例)

SpeakingV / SV (話すための文例)

Vocabulary / V

### 【メタ言語】

Grammar / G (説明文と用例の両方を含む)

## コーパス対象のレベル別冊数

レベル	冊数
A1	7
A2	8
B1	14
B2	15
C1	9
C2	1
合計	54

※2013/03/21時点。概数：30000×54＝162万語程度



## まとめ

- ・Inputデータである教材コーパスを作成し、data-drivenでCEFRレベルごとの言語特徴を抽出する
- ・CEFR-Jへの対応付けを考える
- ・技能別、項目別でタグをつけ、それらがどの程度影響しているかも考慮する