

自然言語処理システム

(2) 機械翻訳

機械翻訳の立場

1. 単語と構造の等価性を重視する立場
2. 意味内容の等価性を重視する立場
3. 効果の等価性を重視する立場

形式的対応のずれ（１）

Ⅰ. 語彙の対応のずれ

ex. (服を)着る、(めがねを)かける、(帽子を)かぶる
put on {clothes, one's glasses, hat, etc.}

飲む スプーンでスープを飲む
drink eat soup with a spoon

食べる
eat

形式的対応のずれ（２）

2. 文法レベルと語彙レベルのずれ

彼は黒いスーツを**着ている**
He **wears** a black suit.

着ている = 着る + ている
(アスペクト: 進行)

3. 単語と句の対応

湯
hot water

形式的対応のずれ (3)

4. 構造のずれ

彼は黒いスーツを着ている	SOV
He wears a black suit.	SVO

5. 品詞間のずれ

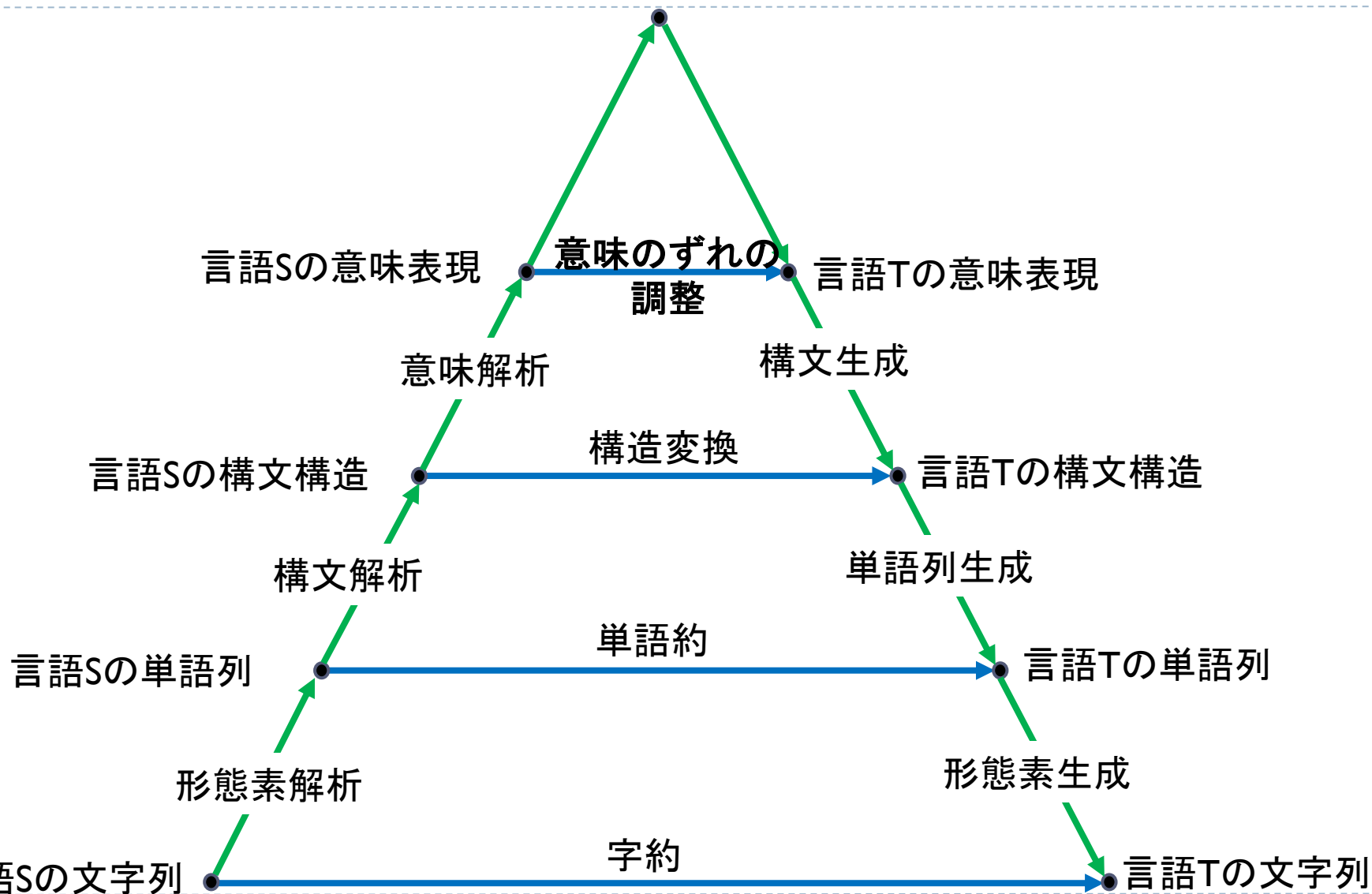
刺激性の気体	名詞 + 「の」
stimulative gass	形容詞

6. 明示的に表現する情報のずれ

英語	数(単数、複数)
日本語	なし

機械翻訳の方式

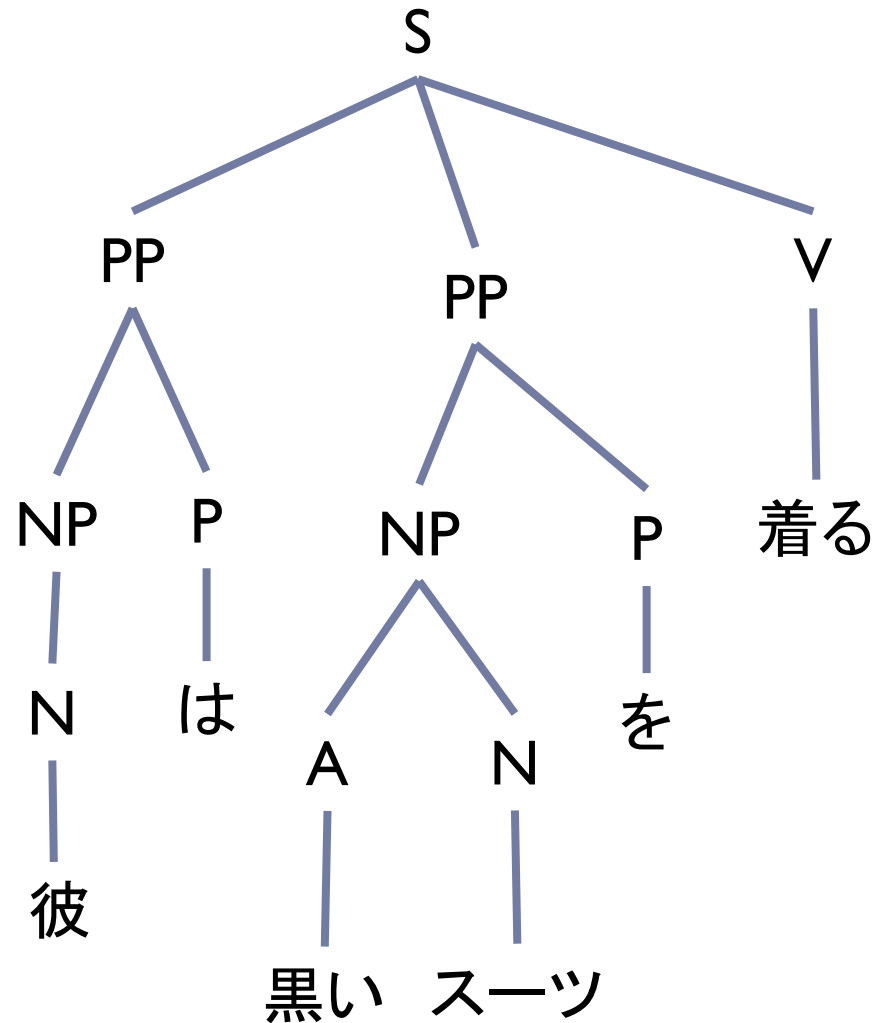
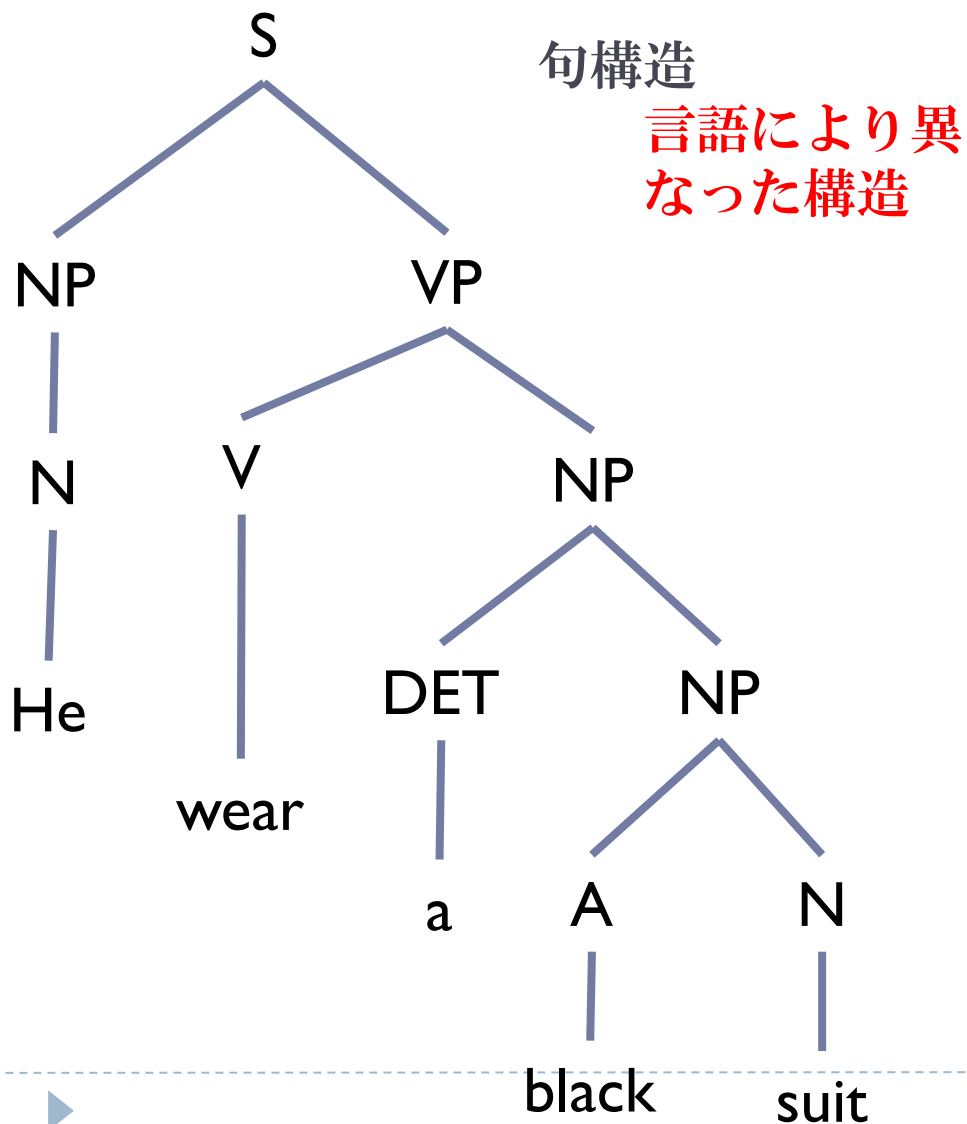
言語に依存しない意味表現(中間言語)



句構造と格構造

He wears a black suit.

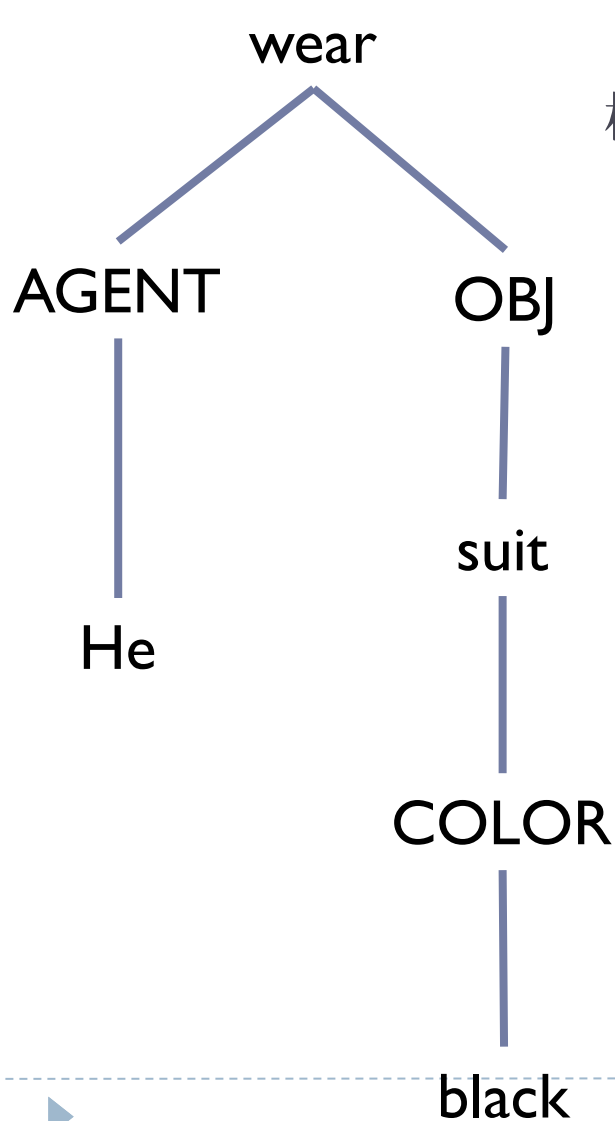
彼は黒いスーツを着ている。



句構造と格構造

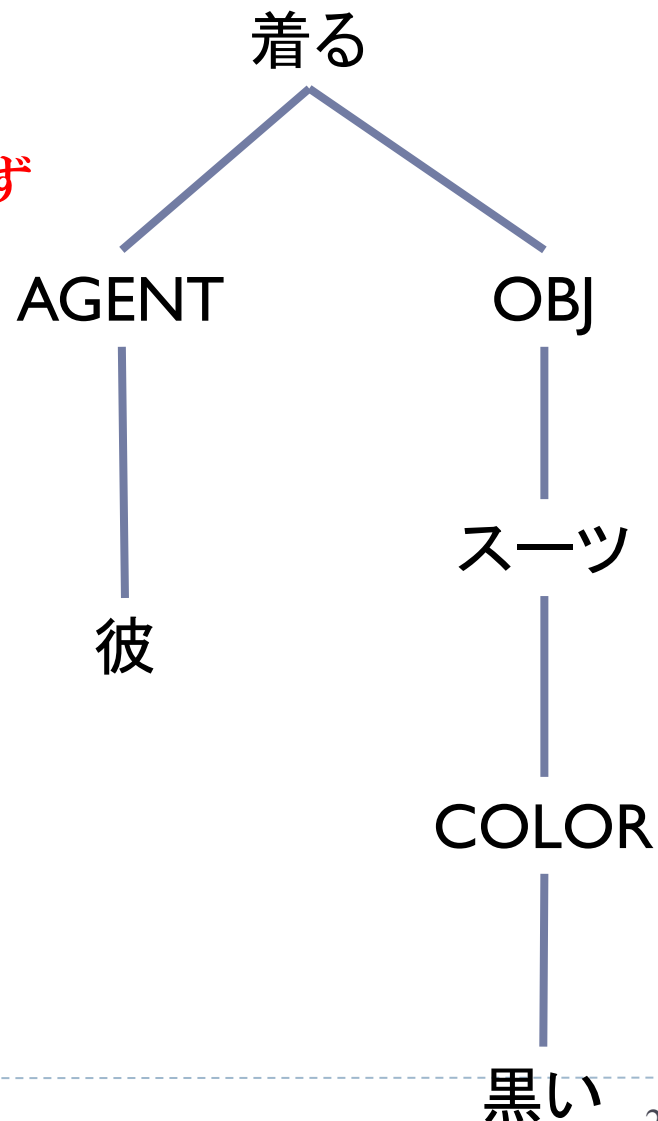
He wears a black suit.

彼は黒いスーツを着ている。



格構造

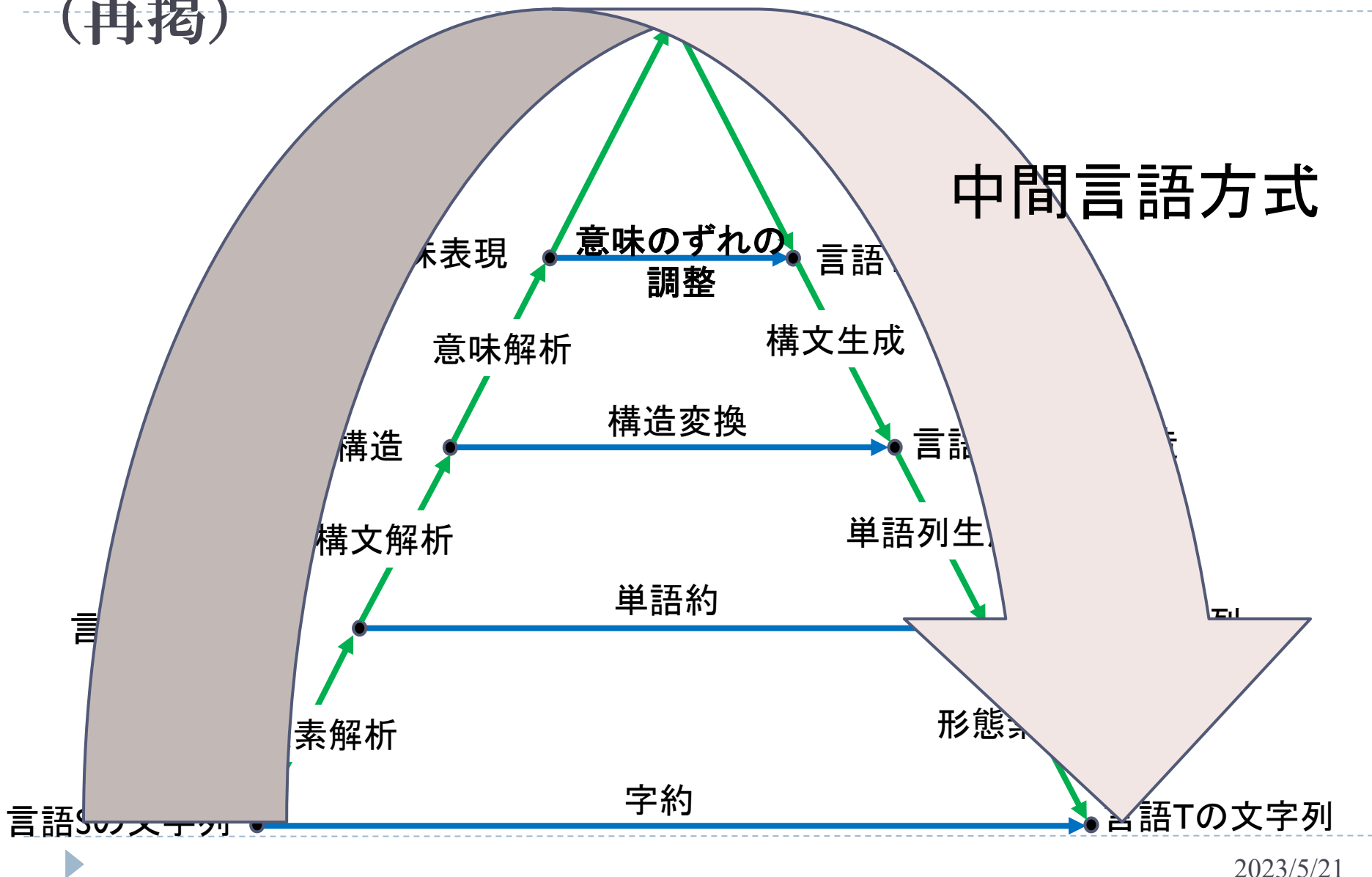
言語によらず
等価な構造



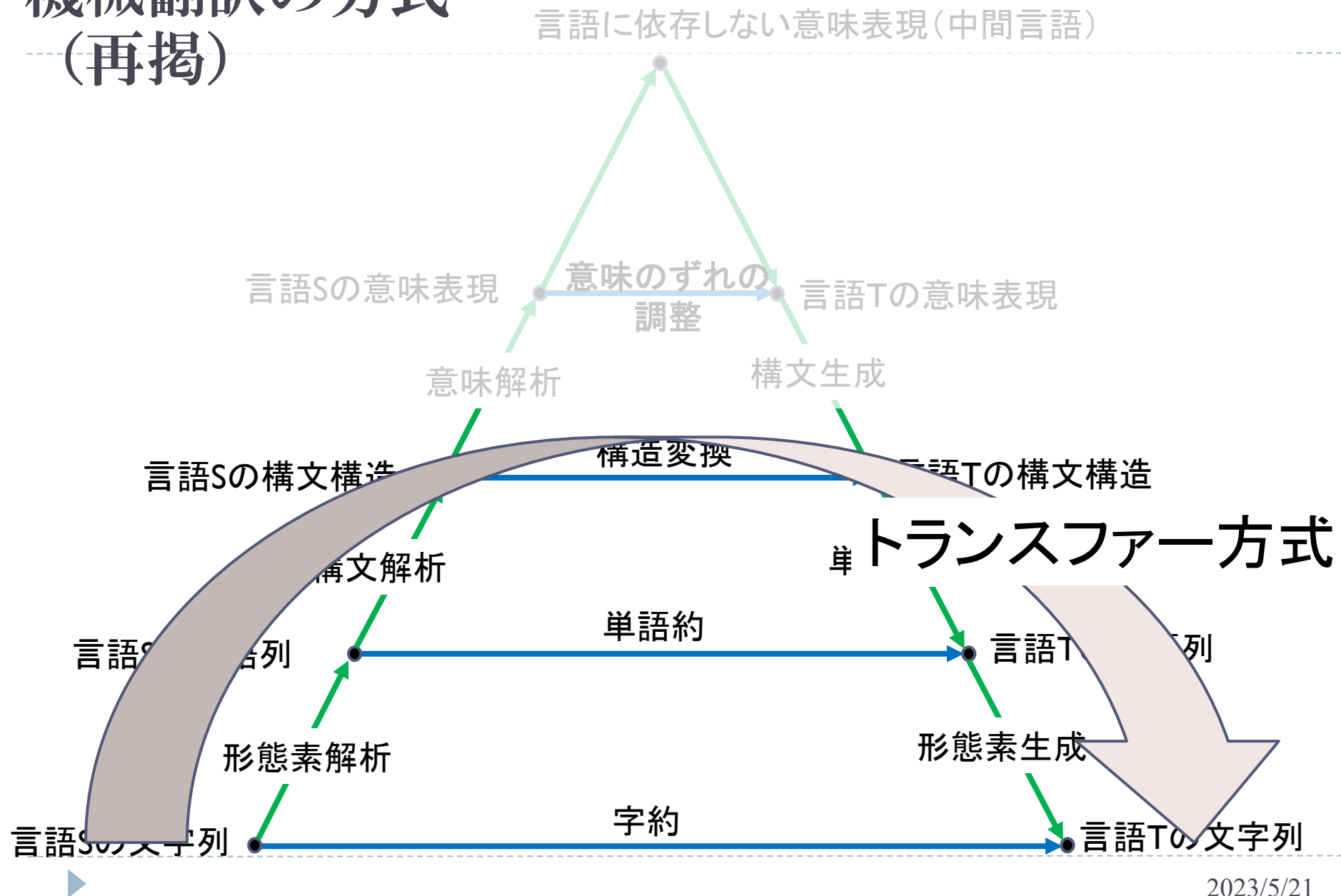
機械翻訳の方式 (再掲)

言語に依存しない意味表現(中間言語)

中間言語方式



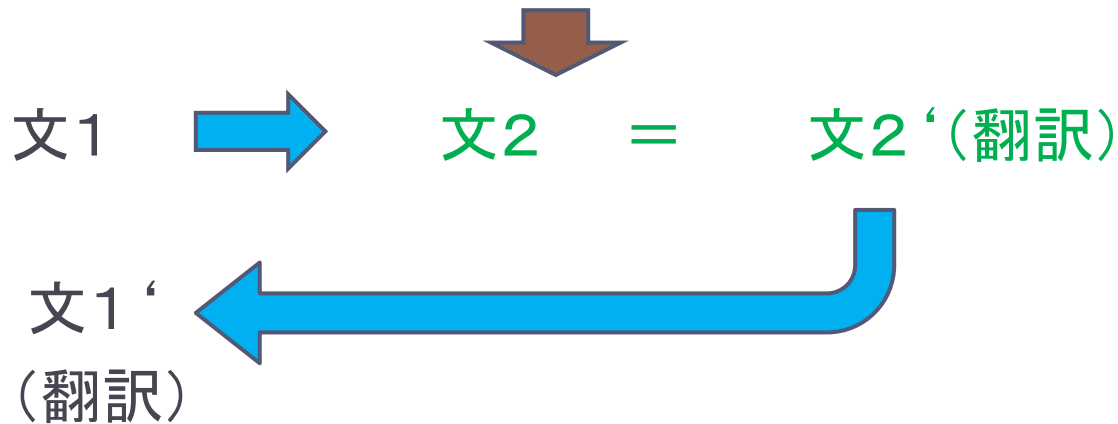
機械翻訳の方式 (再掲)



実例型機械翻訳方式

基本原理

ある文と似た文は、元の文と同じように訳される(ことが多い)



例) Swimming is good for you.

* 任意の大きさのテキスト
(句、節)に適用可能

Vegetable is good for you. = 野菜はからだにいい。

水泳はからだにいい。

「AのB」の実例型機械翻訳

	入力		出力
A	付属語	B	(パターン)
8日	の	午後	B of A
会議	の	参加料	B for A
京都	での	会議	B in A
...

入力: 東京 での 滞在

ニューラル機械翻訳

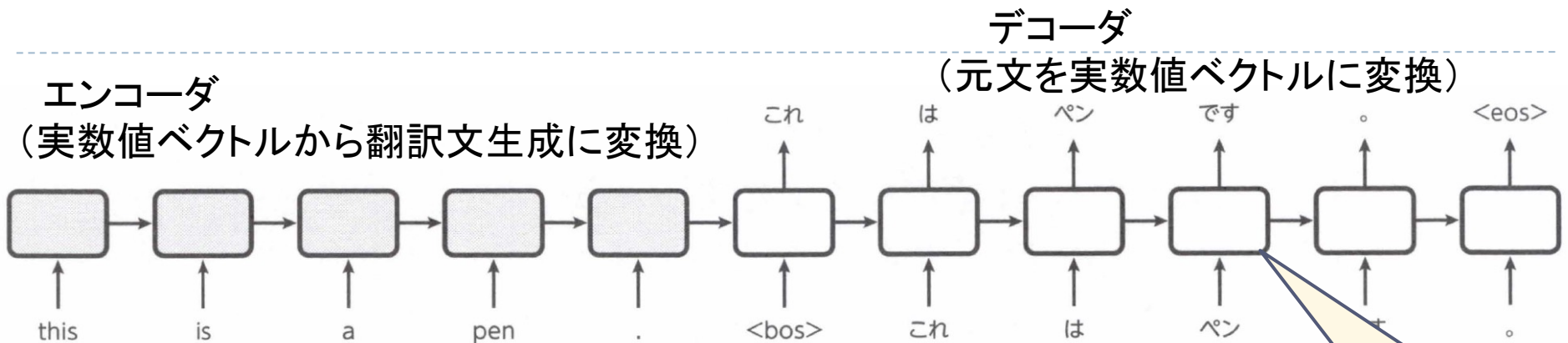


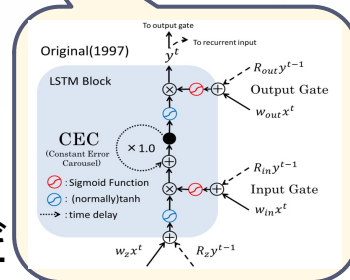
図 6-1 : Seq2Seq (Encoder-Decoder) の概要図

機械学習のための対訳コーパス

they finally acknowledged it as true .
he didn 't care for swimming .
he is no less kind than his sister .
you must be back before ten .
break a leg .
she lives next door to us .
i 'm about to tell you the answer .
.....

彼らはついにそれが真実だ
彼は水泳が得意ではなかった。
彼はお姉さんに劣らず親切だ。
10 時前に戻らなければならない。
成功を祈るわ。
彼女は私たちの隣の家にすんでいる。
あなたに返事をしようとしているところ
です。

.....



実行(epoch:30まで続く)

```
bigMac:6 tam$ python 03_encoder_decoder_tf.py
2020-12-20 22:20:05.345306: I tensorflow/core/platform/cpu_feature_guard.cc:142]
  This TensorFlow binary is optimized with oneAPI Deep Neural Network Library (on
eDNN)to use the following CPU instructions in performance-critical operations:
AVX2 FMA
To enable them in other operations, rebuild TensorFlow with the appropriate comp
iler flags.
2020-12-20 22:20:05.364320: I tensorflow/compiler/xla/service/service.cc:168] XL
A service 0x7f927dd3a210 initialized for platform Host (this does not guarantee
that XLA will be used). Devices:
2020-12-20 22:20:05.364343: I tensorflow/compiler/xla/service/service.cc:176]
StreamExecutor device (0): Host, Default Version
-----
epoch: 1
loss: 3.509, val_loss: 3.5
> at the end of a working day everybody is in a hurry to get home .
= 一 日 の 仕 事 が 終 わ る と 皆 家 路 を 急 ぐ 。 </s>
< 彼 は は は は は は は は は は を を い 。 </s> </s> </s>

> i am as much in the wrong as you are about not writing before this .
= ご 無 沙 汰 は お 互 い さ ま で す 。 </s>
< 彼 は は は は は は は は は は を を を い 。 </s>

> my father boasts of the fact that he has never had a traffic accident .
= 父 は 車 の 無 事 故 を 自 慢 し て い る 。 </s>
< 彼 は は は は は は は は は は を を を い 。 </s>

> mr smith and i have been acquainted with each other for a long time .
= ス ミ ス 私 と は 長 い 間 の 知 り 合 い で す 。 </s>
< 彼 は は は は は は は は は は を を い 。 </s> </s>
```


まとめ

- ▶ 機械翻訳について
 - ▶ 形式対応のずれ
 - ▶ 機械翻訳の方式
 - ▶ 格構造
 - ▶ 実例型機械翻訳方式
 - ▶ ニューラル機械翻訳

課題

- ▶ 形式対応のずれ(1)～(3)の例文1～6(日本語、英語)について、中国語の場合を加えよ。