

句構造へのアテンションに 基づくニューラル機械翻訳モデル

江里口瑛子, 橋本和真, 鶴岡慶雅 (東京大学)

目次

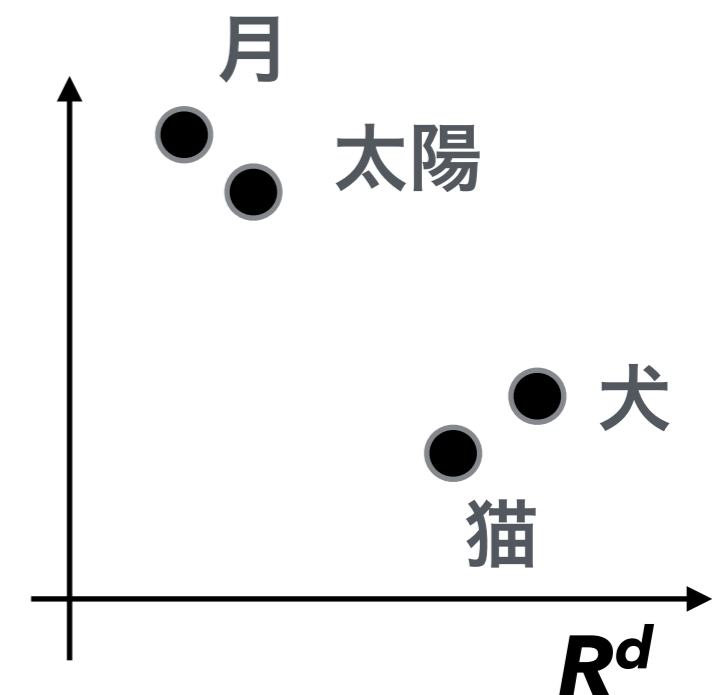
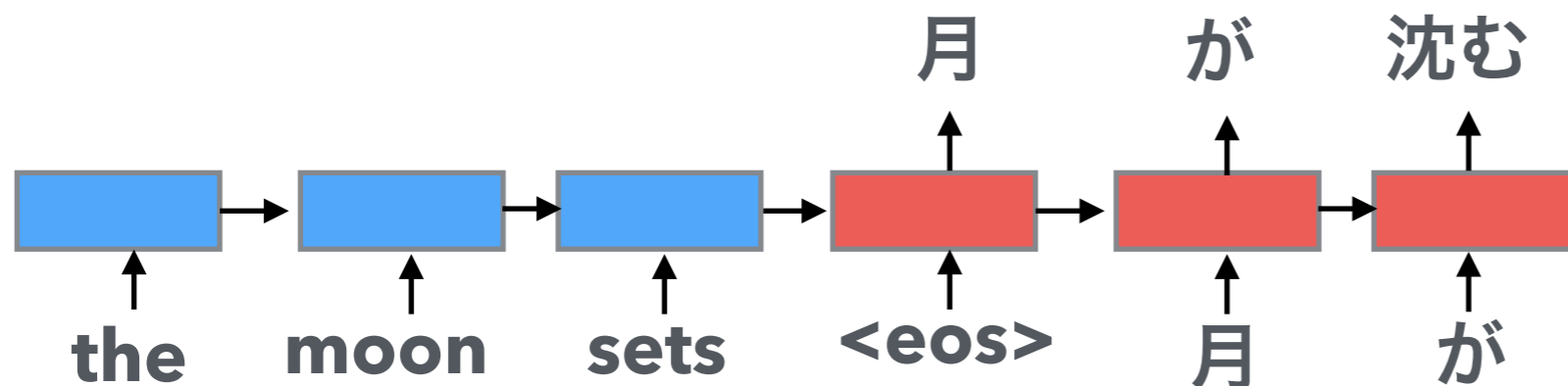
- 背景・目的
- 提案手法
 - 句構造のためのエンコーダ
 - 句構造へのアテンション機構
- 実験
- まとめと今後の課題

目次

- 背景・目的
- 提案手法
 - 句構造のためのエンコーダ
 - 句構造へのアテンション機構
- 実験
- まとめと今後の課題

ニューラル機械翻訳: エンコーダ・デコーダモデル

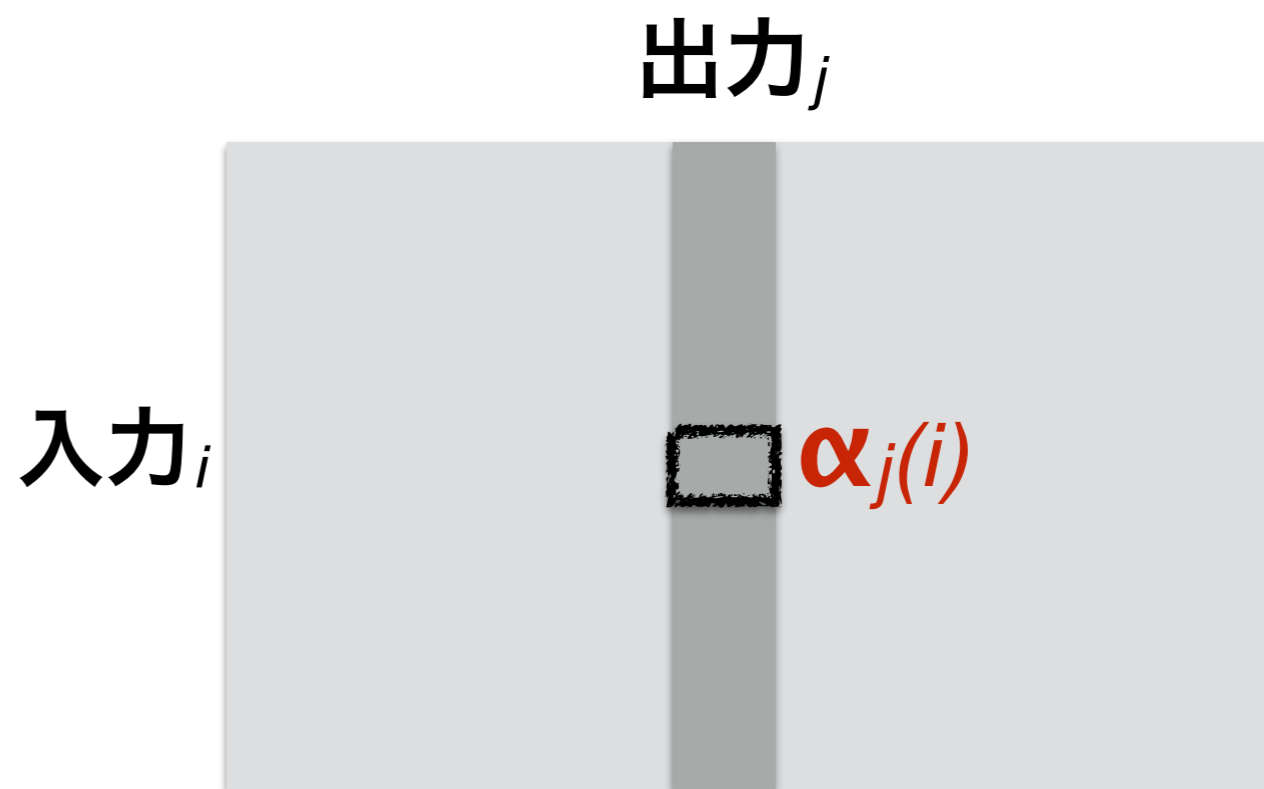
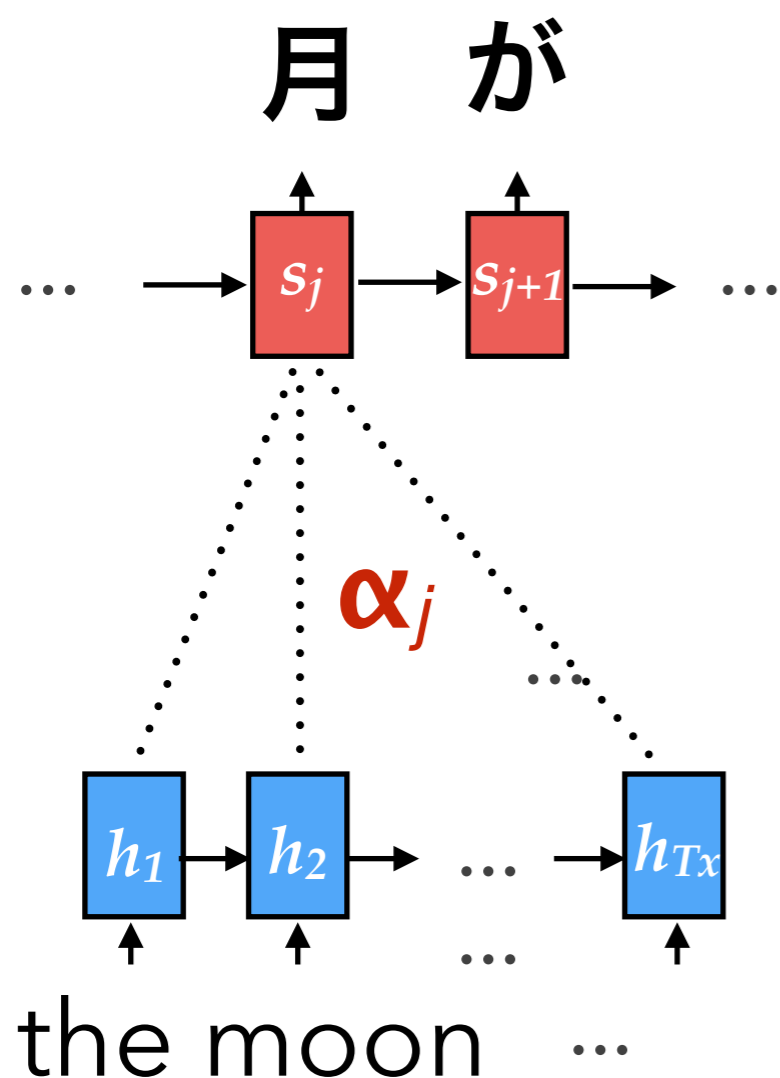
- 単語や文をベクトルで表現する研究が盛ん
- ニューラルネットワークを用いた翻訳モデル
 - メモリユニット (LSTM) による可変長系列変換
(Sutskever et al., 2015)
 - 大規模対訳コーパスから翻訳モデル $p(\mathbf{y}|\mathbf{x})$ を最大化
 - 一部言語対 (英仏・英独) でSOTA
(Luong et al., 2015a; Luong et al., 2015b)



系列へのアテンションに基づくモデル (ANMT)

(Bahdanau et al., 2015; Luong et al., 2015b)

- デコード時、各エンコーダ h_i への注目度合い $\alpha_j(i)$ を同時学習
- 入出力の**対応関係**を学習
- 長文・固有名詞の翻訳



本発表

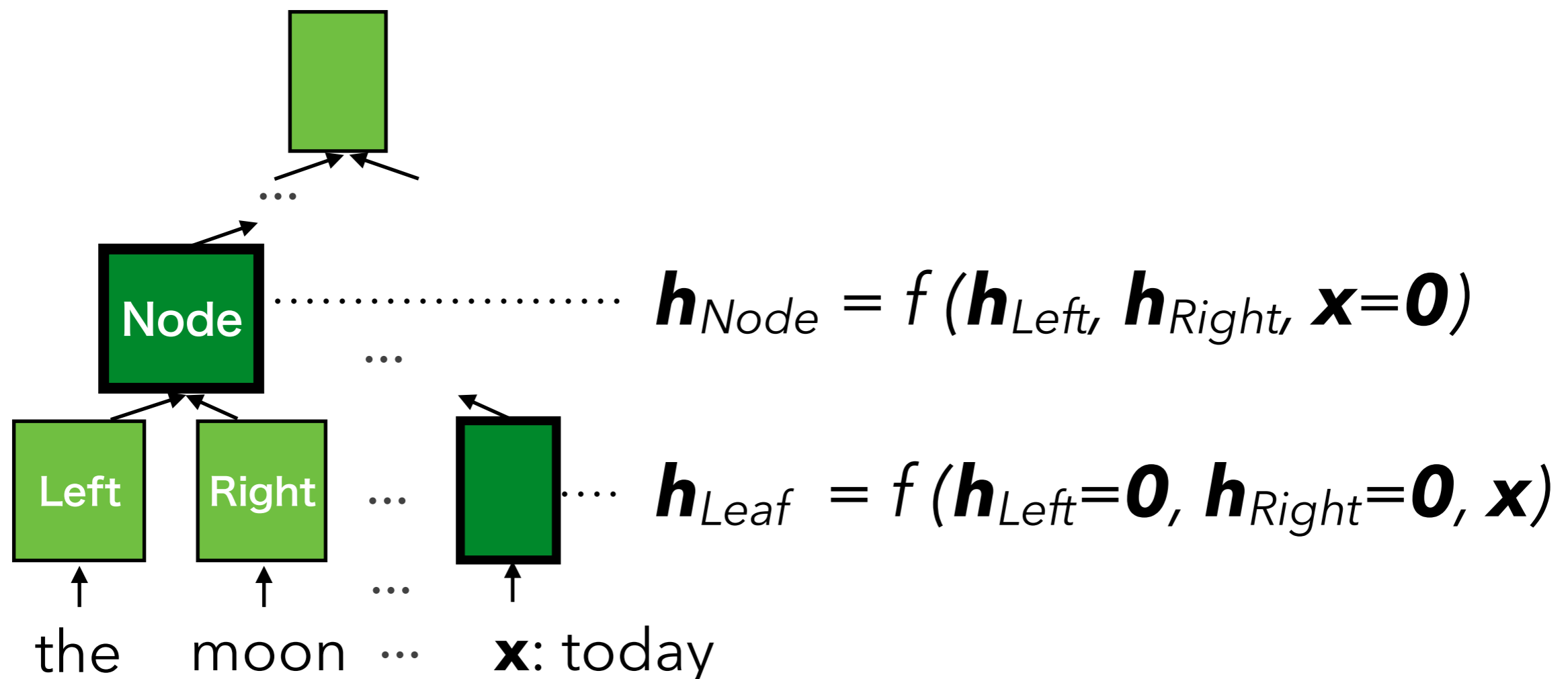
- 構文情報 (句構造) を利用したNMTモデルの提案
 - 言語によって、事前並び替えや構文構造考慮([Neubig, 2013](#))が有効
 - NMT: 逆順入力 ([Sutskever et al., 2014](#)), 構造ゲート ([Cho et al., 2014](#))
 - 入力文の句構造に従ったエンコーダの作成
- 句構造へのアテンション機構を導入
 - アテンションの補強・範囲限定 ([Cohn et al., 2016](#); [Luong et al., 2015b](#))
 - 最適アテンション窓幅 → 「句/ 単語の単位」
 - 入力句/ 単語と出力単語の対応の同時学習モデル

目次

- 背景・目的
- 提案手法
 - 句構造のためのエンコーダ
 - 句構造へのアテンション機構
- 実験
- まとめと今後の課題

句構造のためのエンコーダ

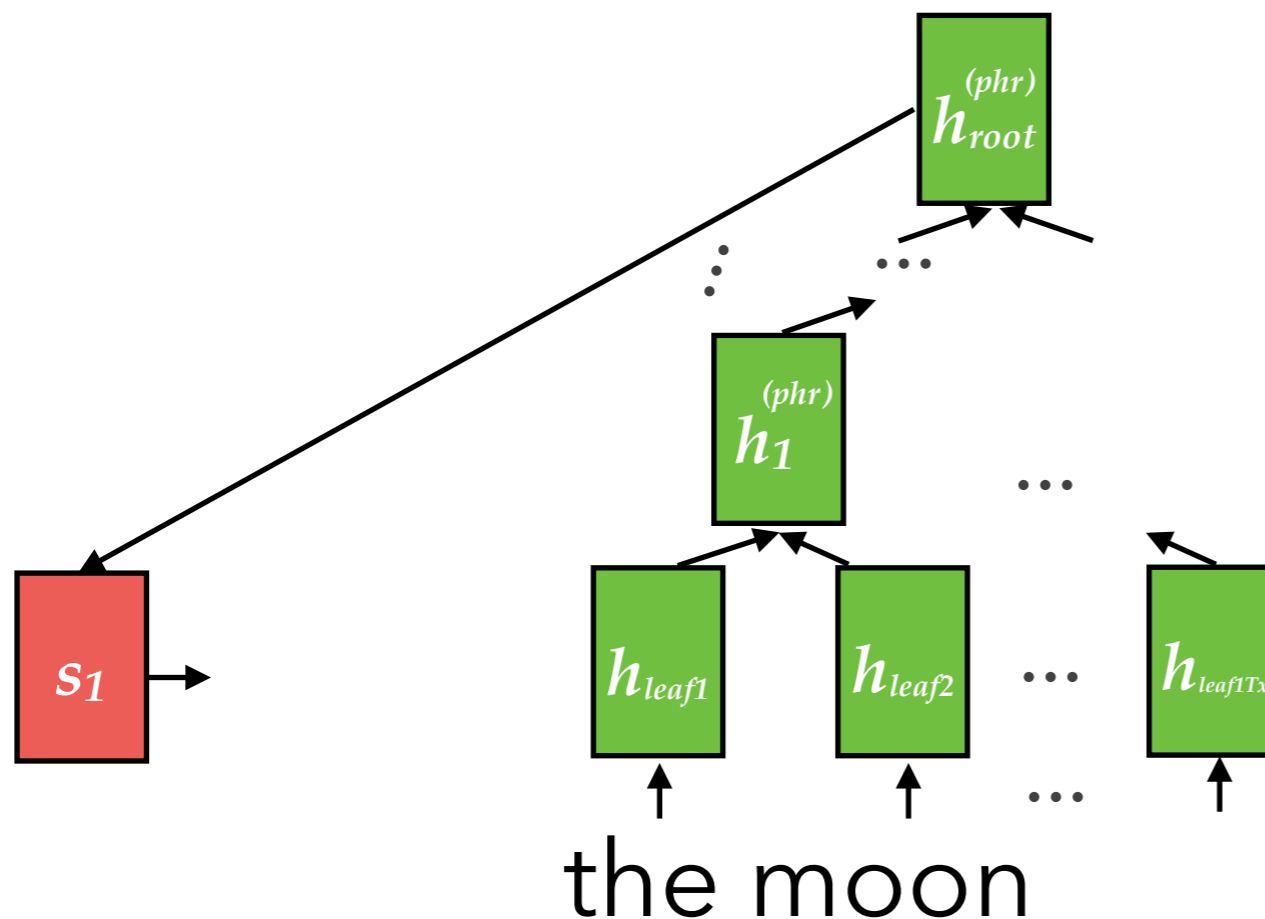
- 文は単語の系列ベクトルではなく、句ベクトルから構成
- 入力文の句構造 (2分木) に従い、句の隠れ層 h を算出
(Tai et al., 2015)



句構造へのアテンション機構

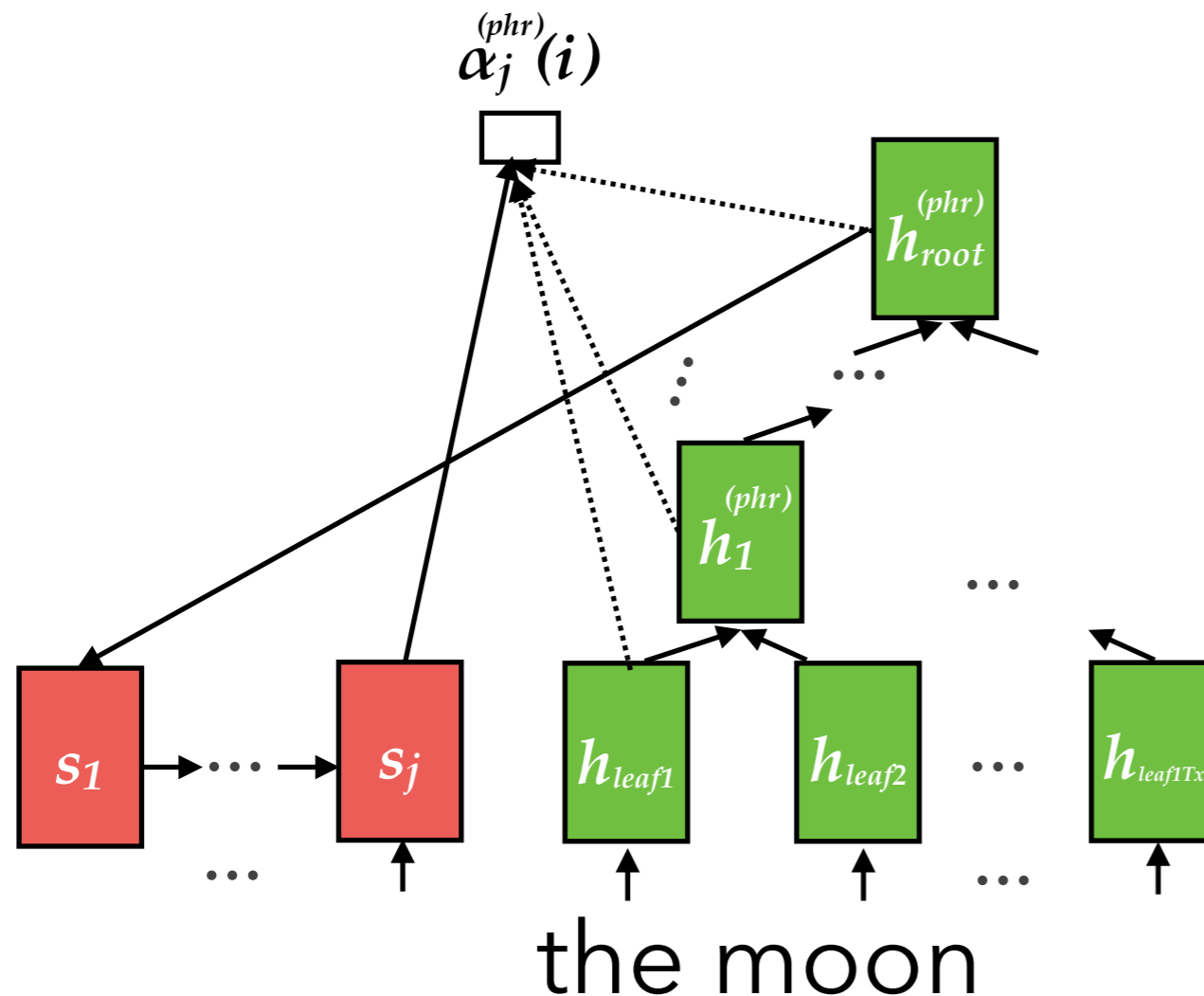
- 句および単語 (葉) へのアテンション機構

- 最終層をデコーダに設定



句構造へのアテンション機構

- 句および単語 (葉) へのアテンション機構

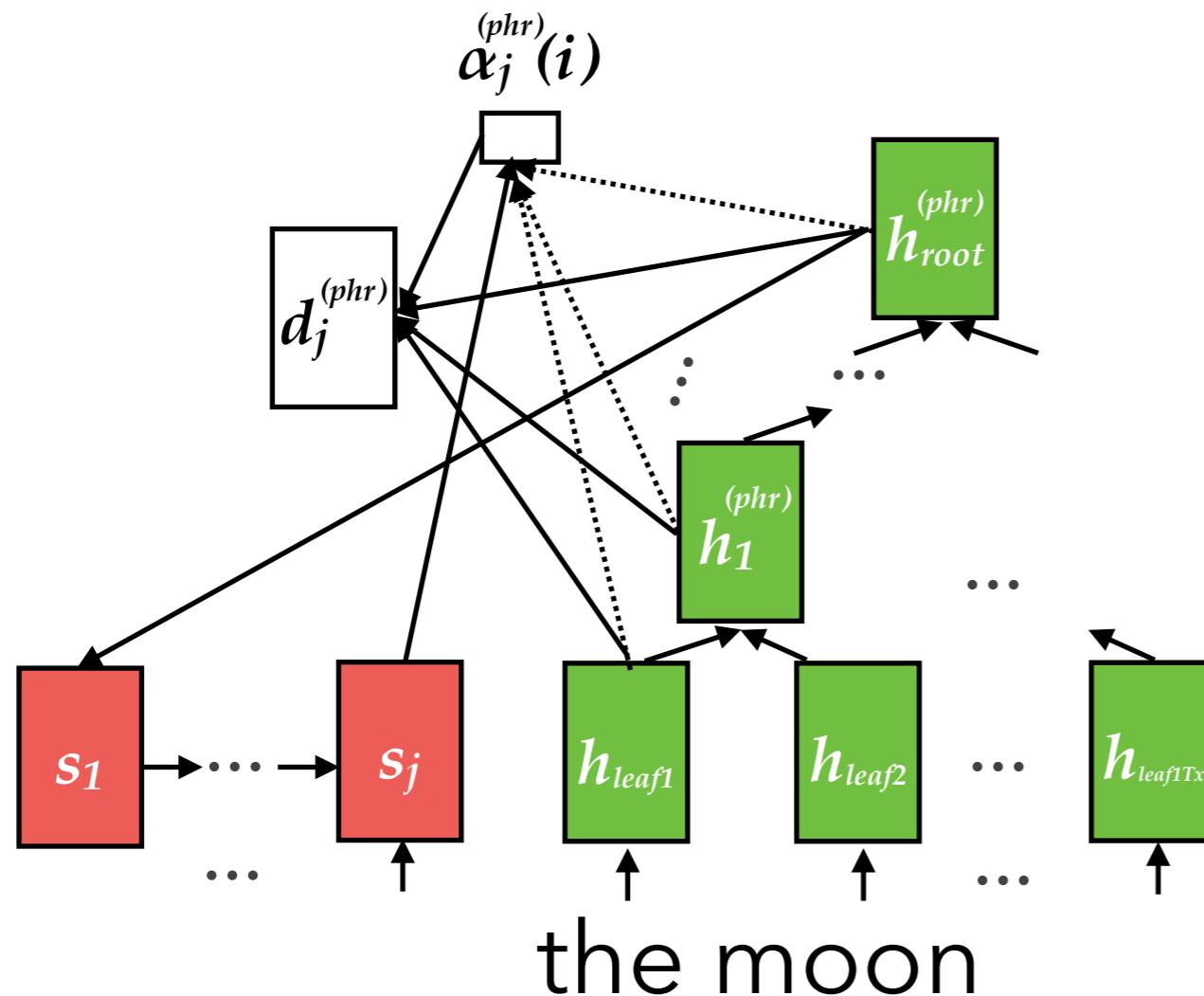


- 最終層をデコーダに設定
- 注目度合 $\alpha_j(i)$ の算出

$$\alpha_j^{(phr)}(i) = \frac{\exp(\mathbf{h}_i^{(phr)} \cdot \mathbf{s}_j)}{\sum_k T_x^{(phr)} \exp(\mathbf{h}_k^{(phr)} \cdot \mathbf{s}_j)}$$

句構造へのアテンション機構

- 句および単語 (葉) へのアテンション機構

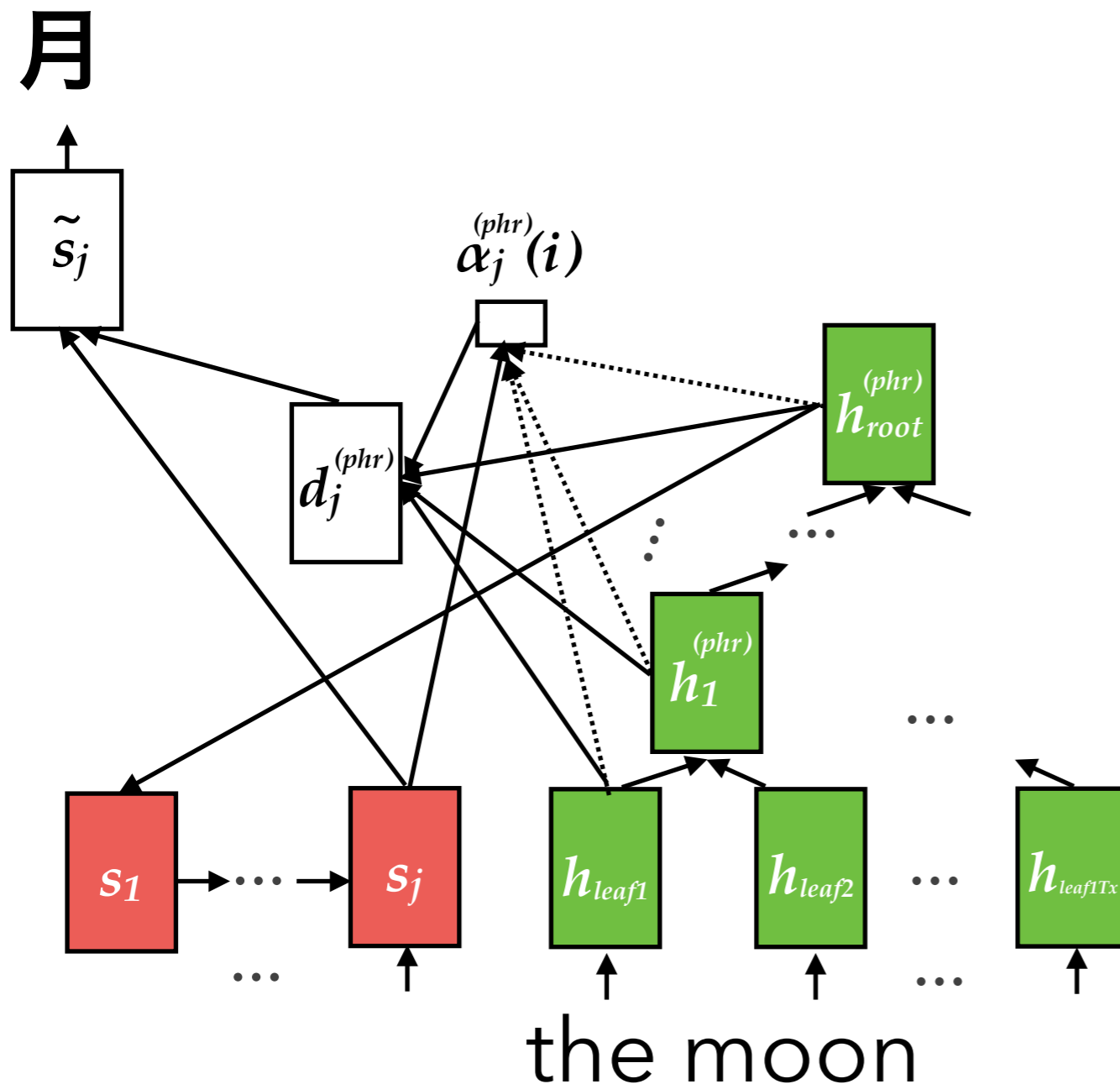


- 最終層をデコーダに設定
- 注目度合 $\alpha_j(i)$ の算出
- エンコーダを重み付け

$$d_j^{(phr)} = \sum_{i=1}^{T_x^{(phr)}} \alpha_j^{(phr)}(i) h_i^{(phr)}$$

句構造へのアテンション機構

- 句および単語 (葉) へのアテンション機構

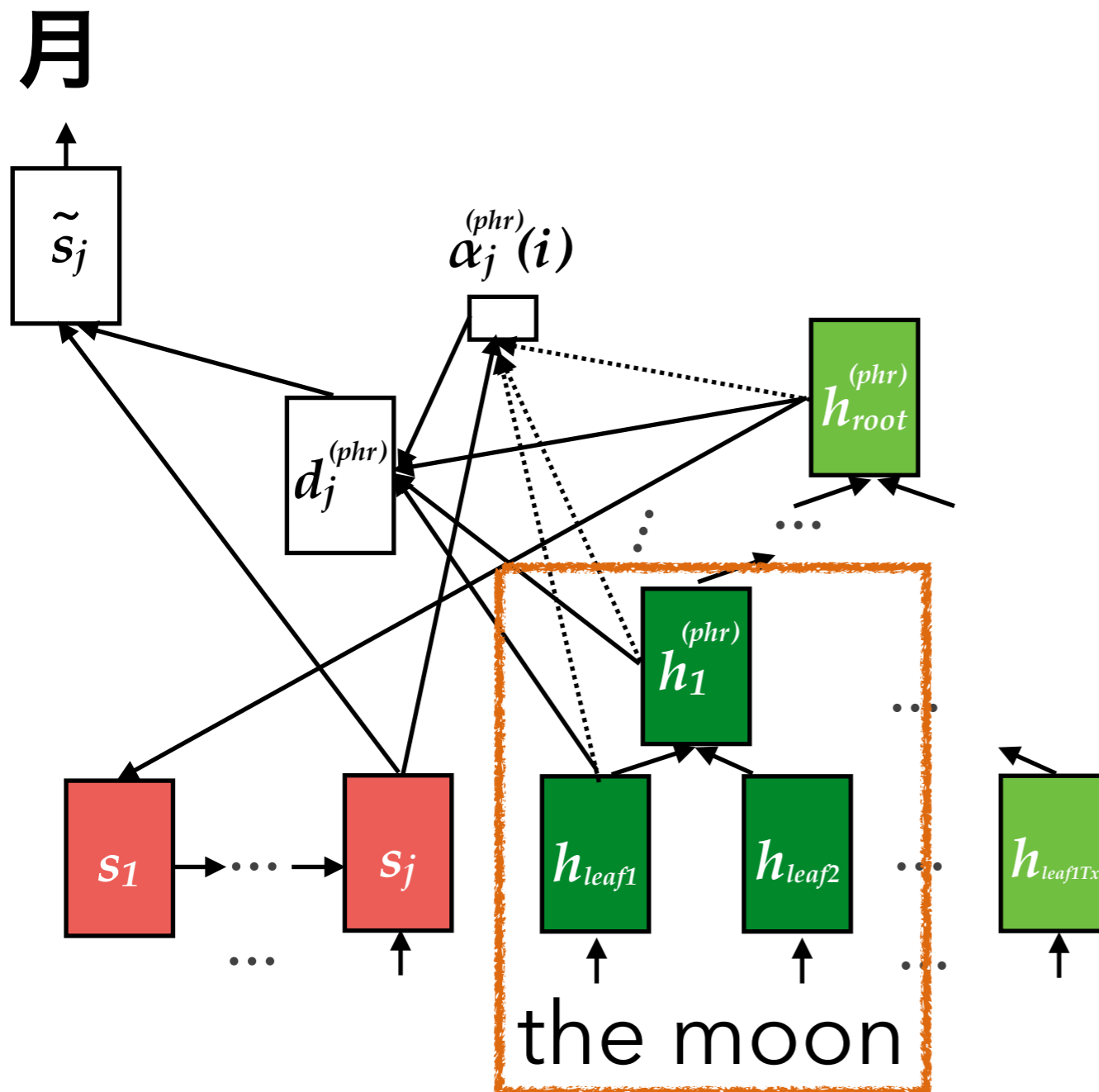


- 最終層をデコーダに設定
- 注目度合 $\alpha_j(i)$ の算出
- エンコーダを重み付け
- 単語予測用デコーダ

$$\tilde{s}_j = \tanh(\mathbf{W}_c[\mathbf{s}_j; \mathbf{d}_j^{(phr)}] + \mathbf{b}_c)$$

句構造へのアテンション機構

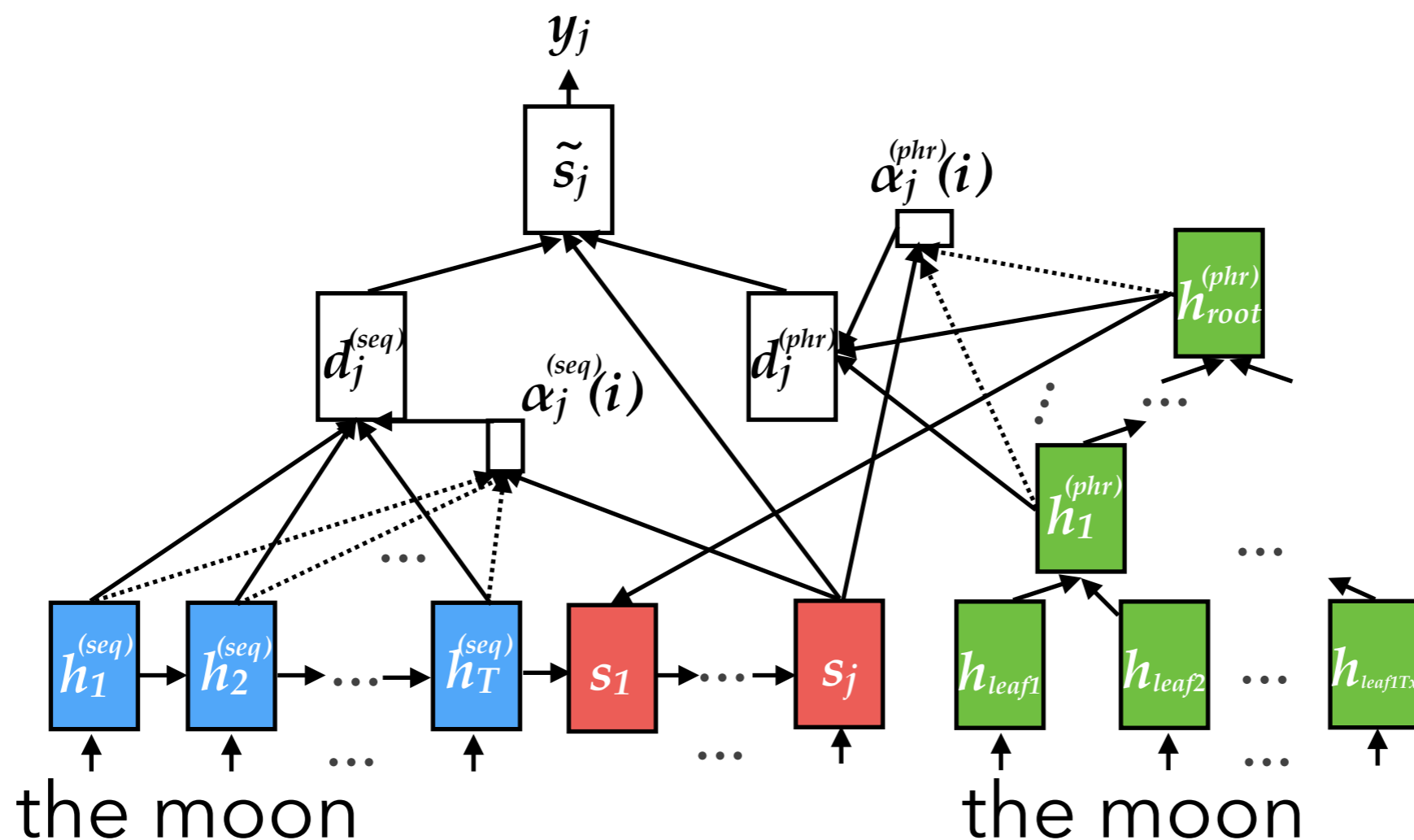
- 句および単語 (葉) へのアテンション機構



- 最終層をデコーダに設定
- 注目度合 $\alpha_j(i)$ の算出
- エンコーダを重み付け
- 単語予測用デコーダ
- 句構造に基づく最適範囲

系列と句への同時注目学習モデル

- 単語出力用デコーダ: $\tilde{s}_j = \tanh(W_e[s_j; d_j^{(seq)}; d_j^{(phr)}] + b_e)$
- 句エンコーダの単語 (葉) はアテンション範囲外
- 2種類のエンコーダの最終層を合成 (Zoph and Knight, 2016)



目次

- 背景・目的
- 提案手法
 - 句構造のためのエンコーダ
 - 句構造へのアテンション機構
- 実験
- まとめと今後の課題

英日翻訳の実験設定

- ASPEC (科学技術論文系の対訳コーパス)
 - 学習データ: 23万文対 (≤ 20 単語); 開発データ: 1,779文; 構文解析成功文
 - 語彙数: (英, 日) = (94K, 60K); Enju (Miyao and Tsujii, 2008), MeCab
- 実験設定
 - 200次元, 50ミニバッチ学習
 - パラメータ更新: 確率的勾配降下法 (スケジュール (Luong et al, 2015b))
 - 勾配クリップ (Sutskever et al, 2015): 5
 - BlackOut (負例サンプリングによるSoftmax近似) (Ji et al, 2016)
 - 出力: ビーム探索 (ビーム幅: 50)
 - C++による実装、マルチコアCPU使用

実験結果

モデル	平均対数尤度
(a-1) ANMT	-117.8
(a-2) ANMT (逆順)	-121.7
(b) 提案手法 (句)	-109.7
(c-1) 提案手法 (系列+句)	-115.0
(c-2) 提案手法 (系列+句; 逆順)	-116.0

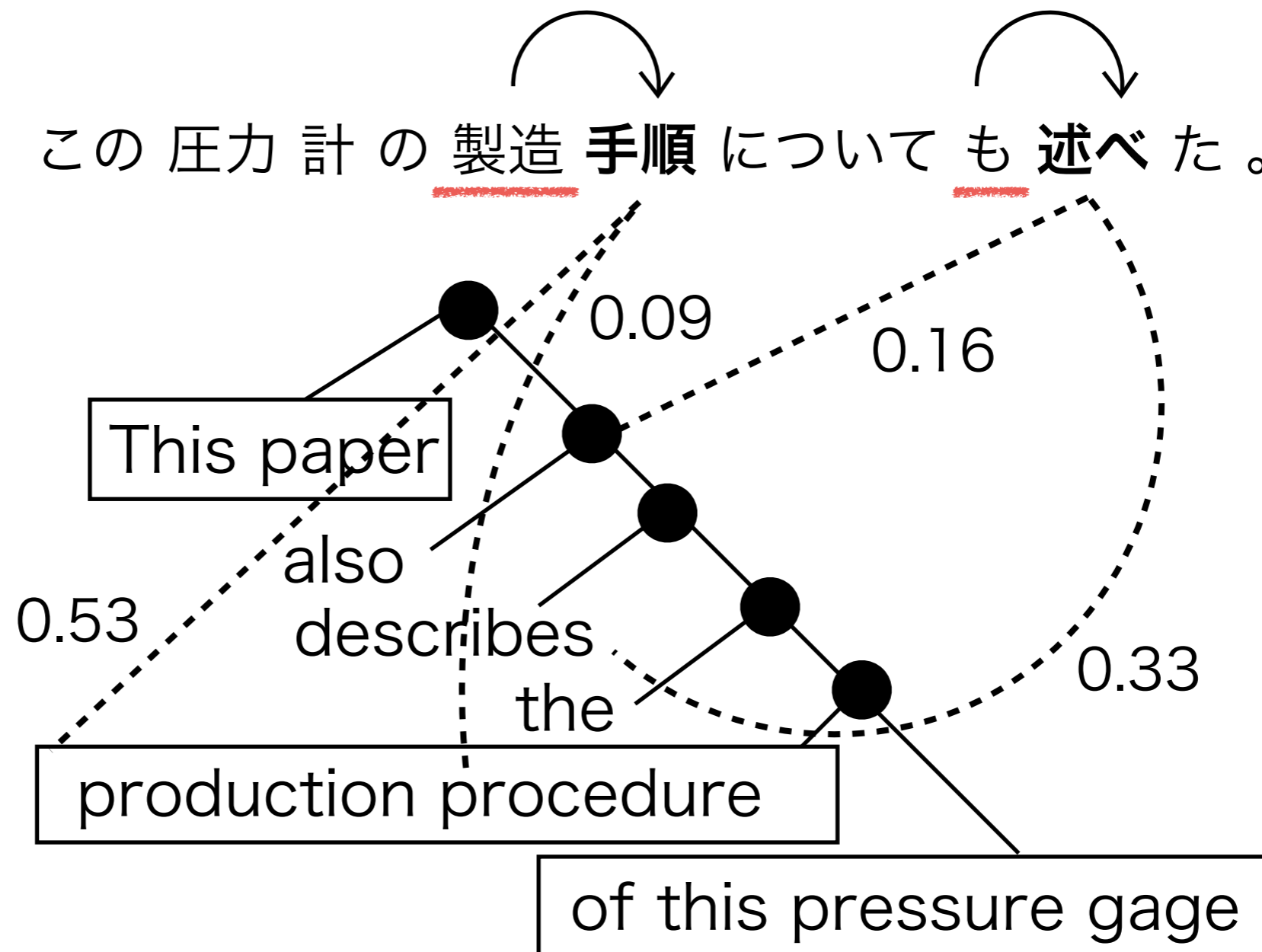
- 入力逆順 (Sutskever et al., 2014) は必ずしも有効でない
- 句構造を利用することでモデルの性能は改善
- 特に、句構造のみに注目した提案手法が最良

翻訳結果の例1

参照訳: 圧力計の製作順序も述べた

出力: この圧力計の製造手順についても述べた。

入力:



翻訳結果の例2

入力: —— [Enju] ——>

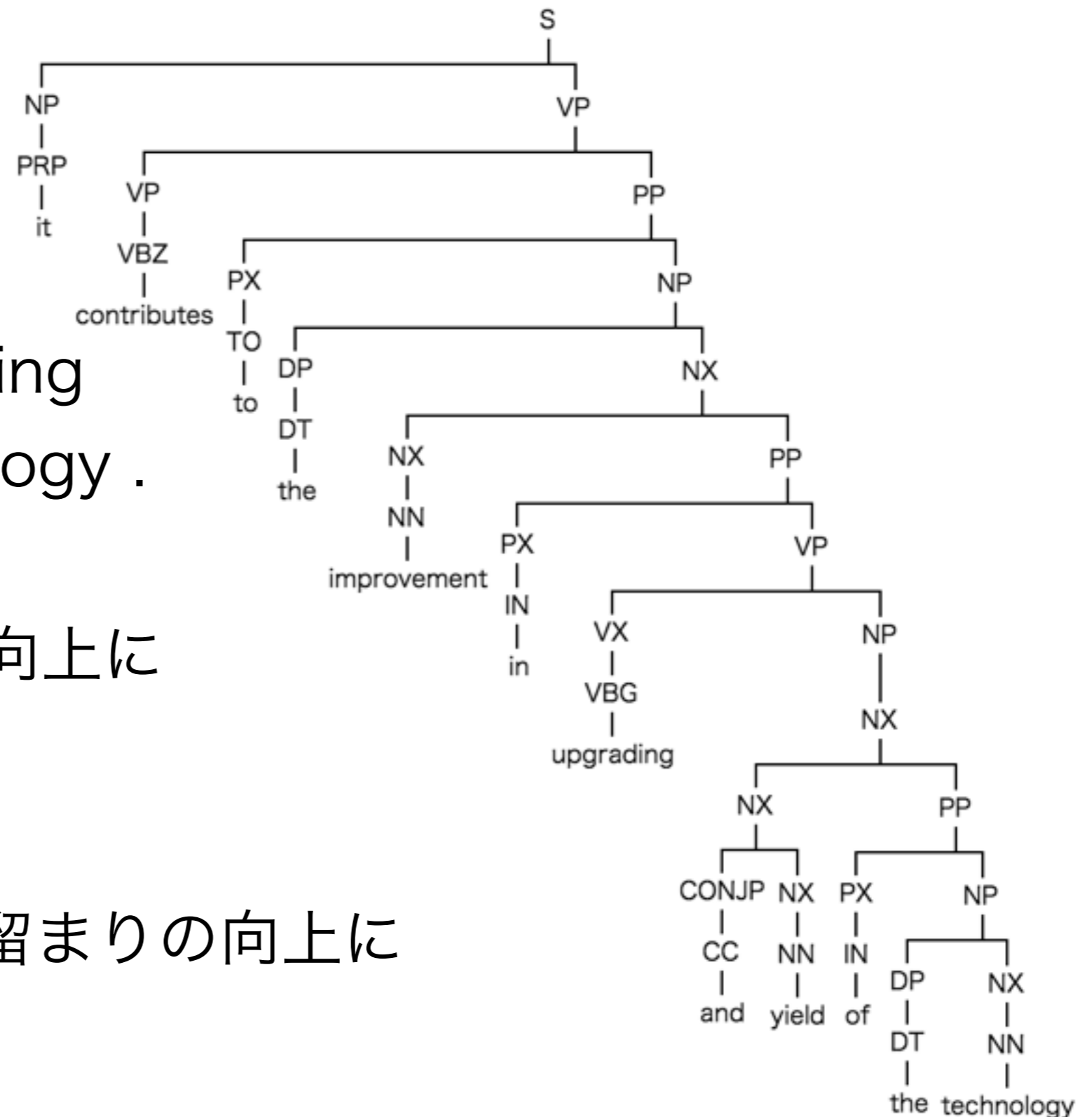
it contributes to the improvement in upgrading and yield of the technology .

出力:

技術の高度化および収量向上に貢献している。

参照訳:

技術のレベルアップと歩留まりの向上に貢献している。



翻訳結果の例2

0.14

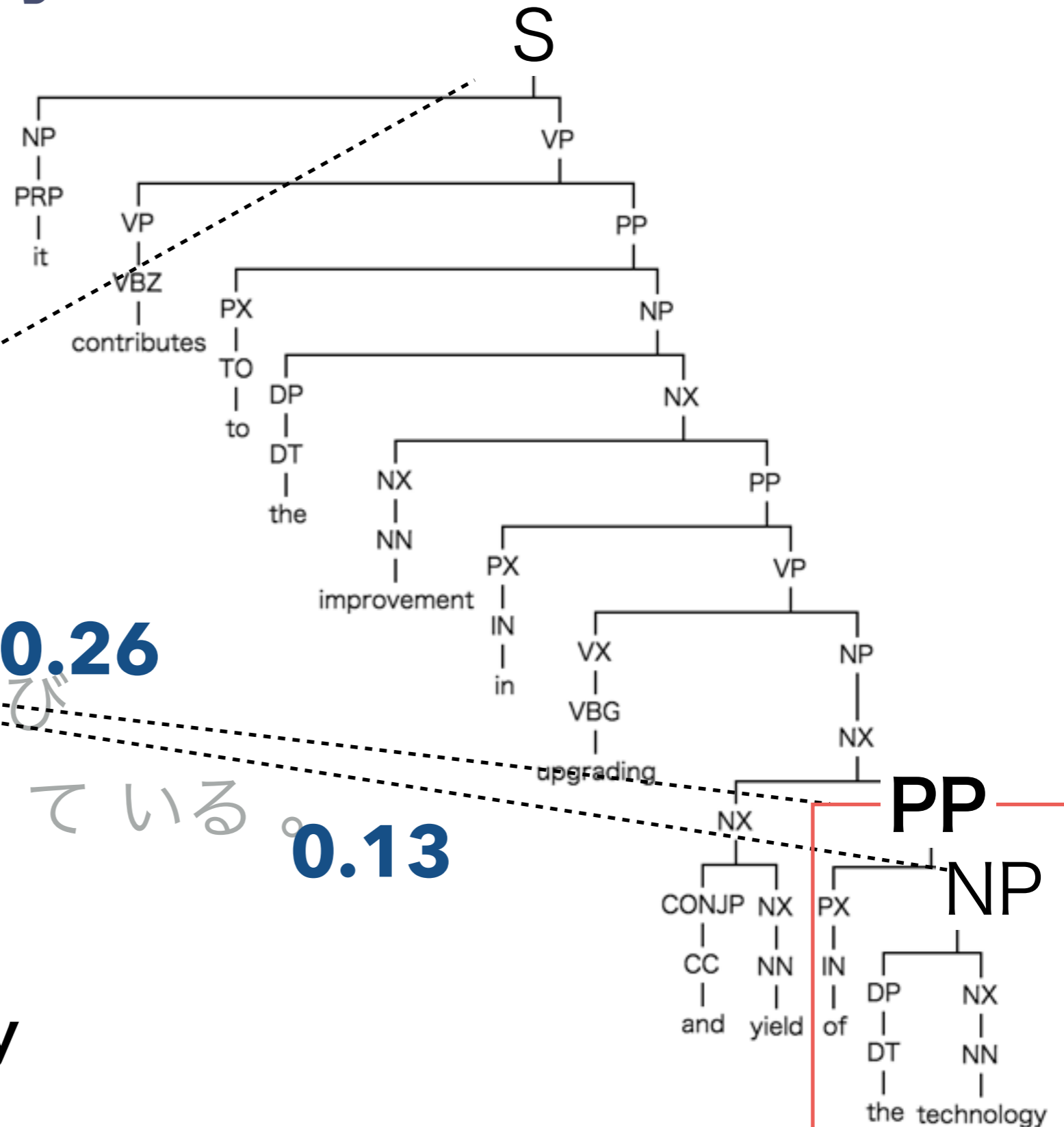
0.26

0.13

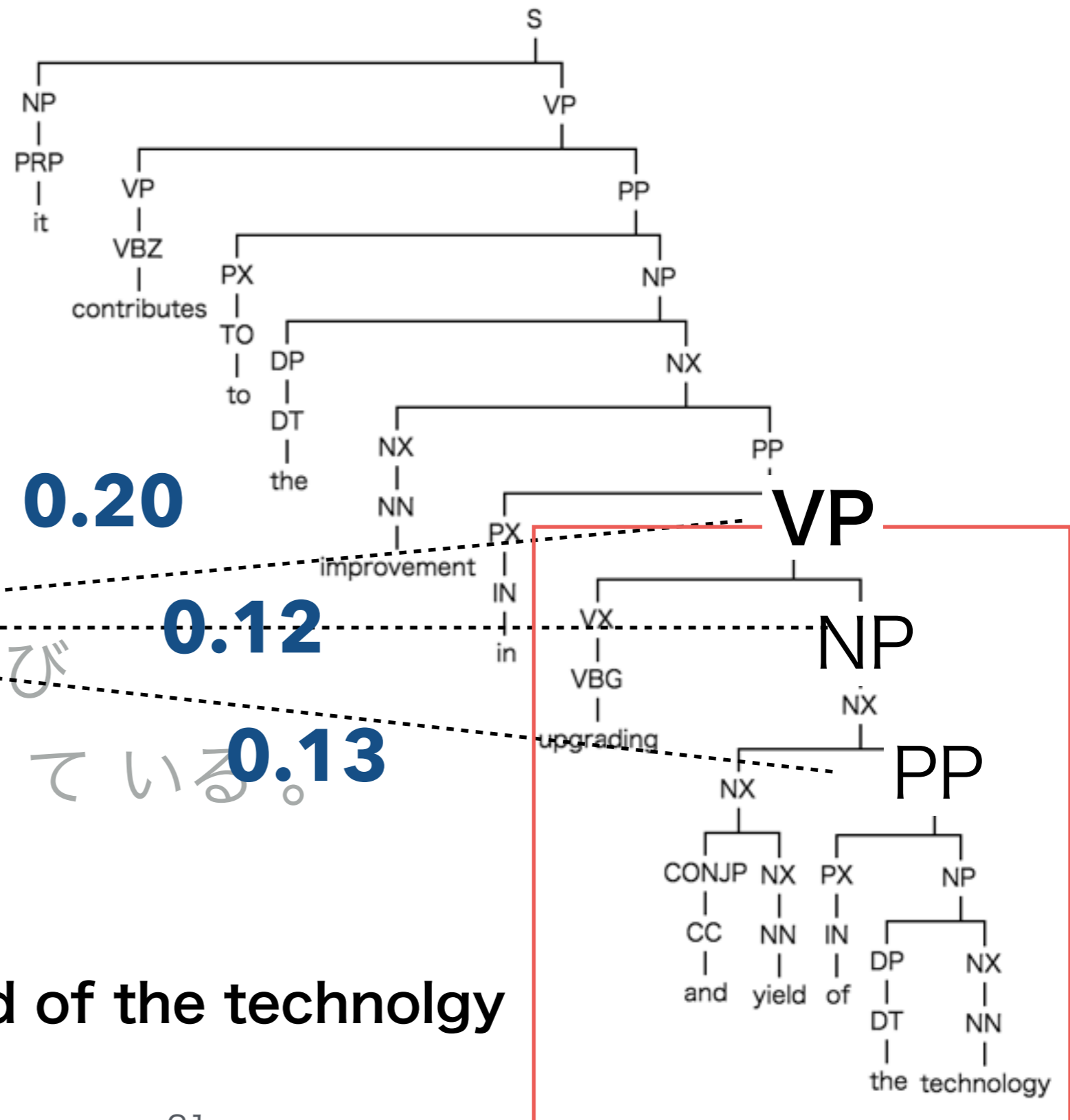
技術の高度化および
収量向上に貢献している。

最注目箇所:

[PP] of the technology



翻訳結果の例2



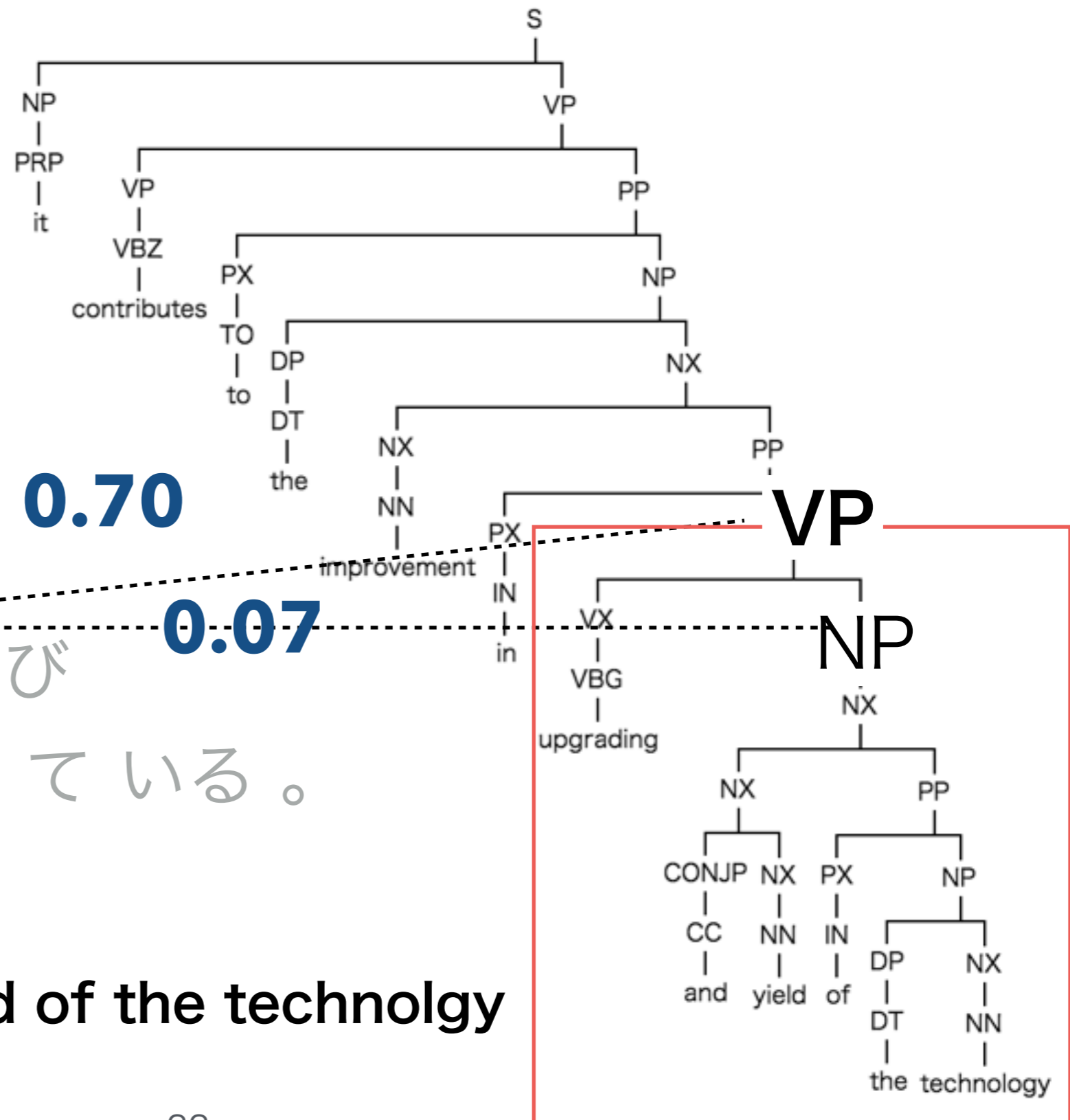
技術の高度化および

収量向上に貢献している。

最注目箇所:

[VP] upgrading and yield of the technology

翻訳結果の例2

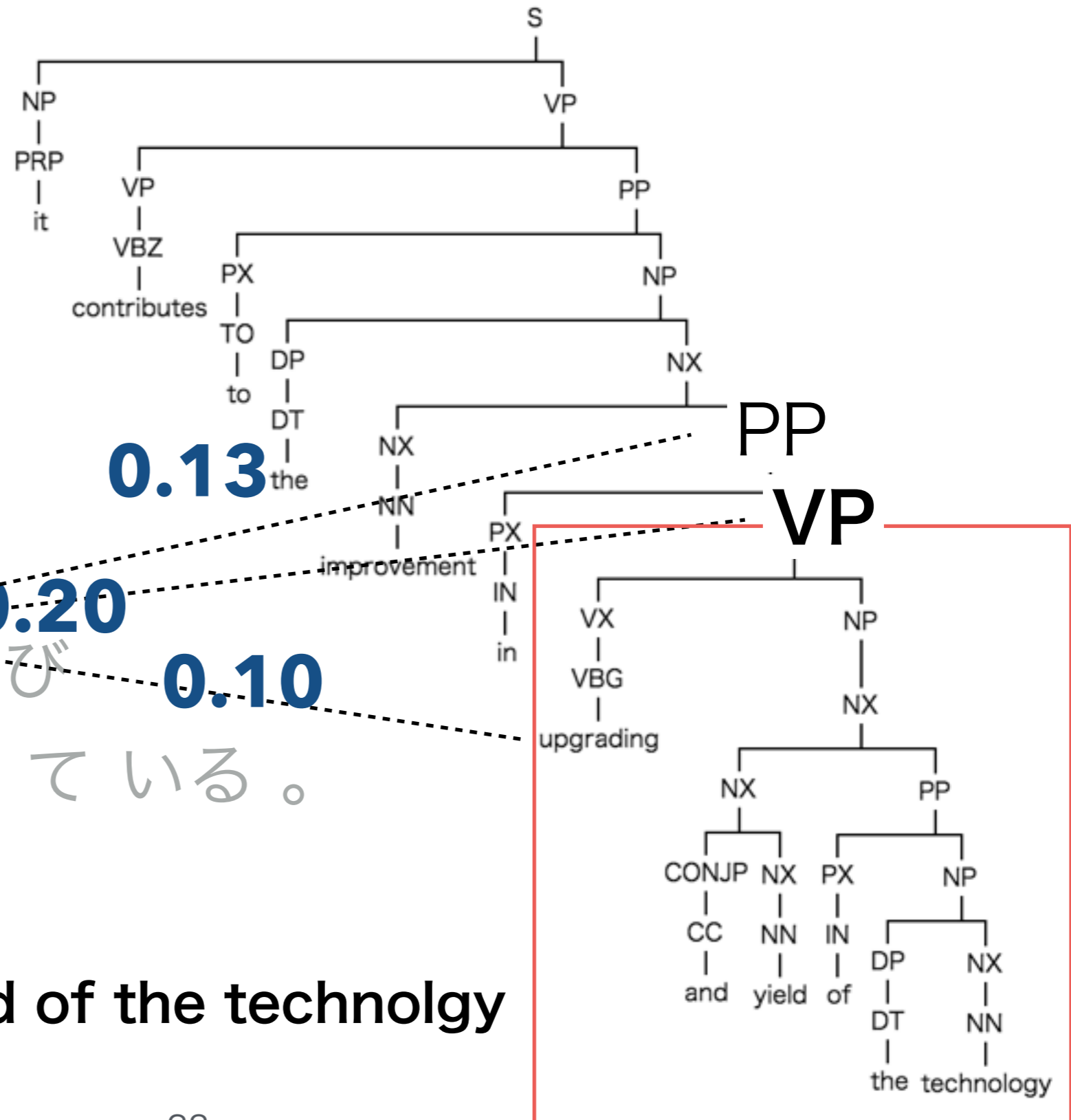


技術の高度化および
収量向上に貢献している。

最注目箇所:

[VP] upgrading and yield of the technology

翻訳結果の例2



0.13

0.20

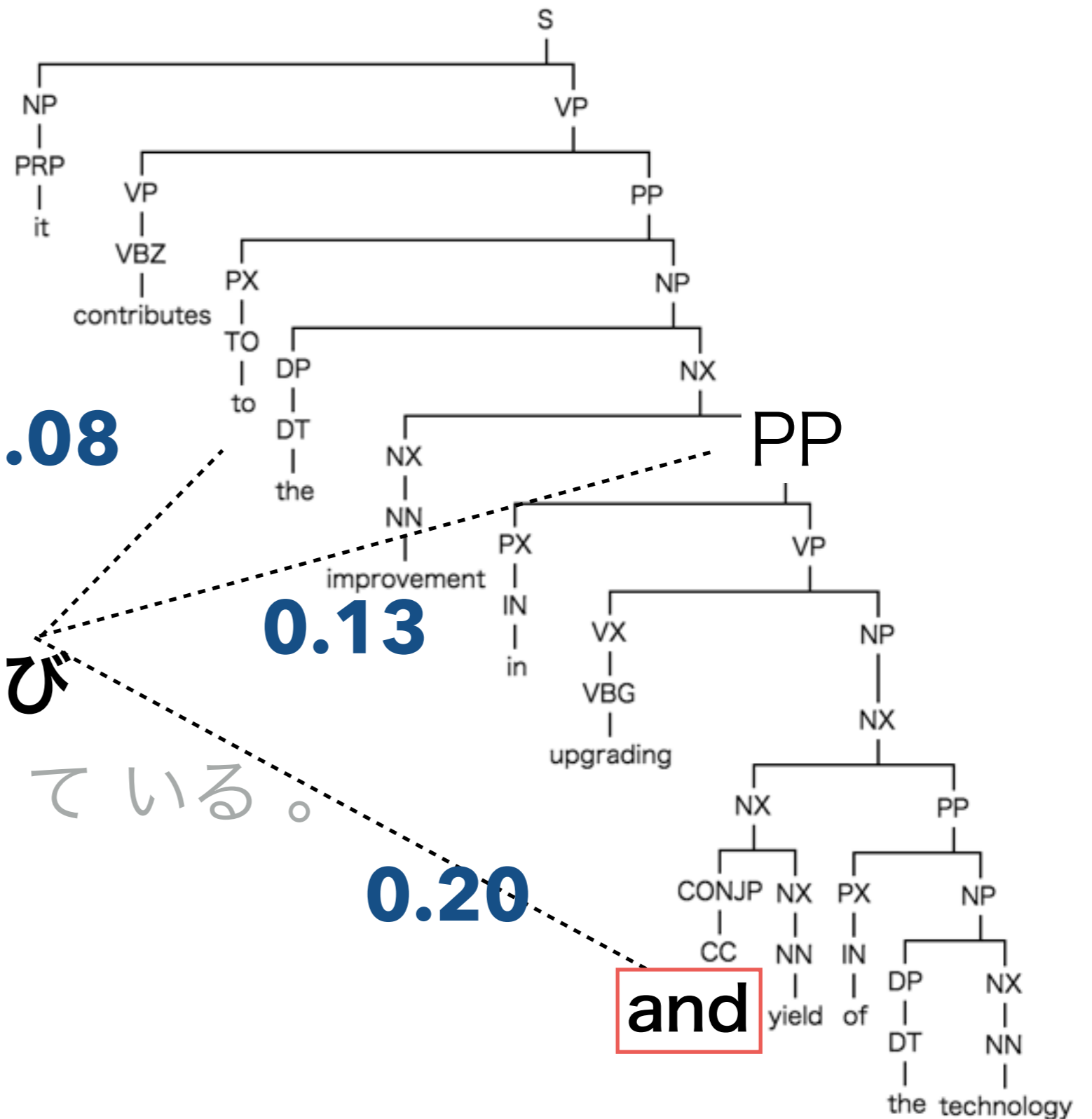
0.10

技術の高度化および
収量向上に貢献している。

最注目箇所:

[VP] upgrading and yield of the technology

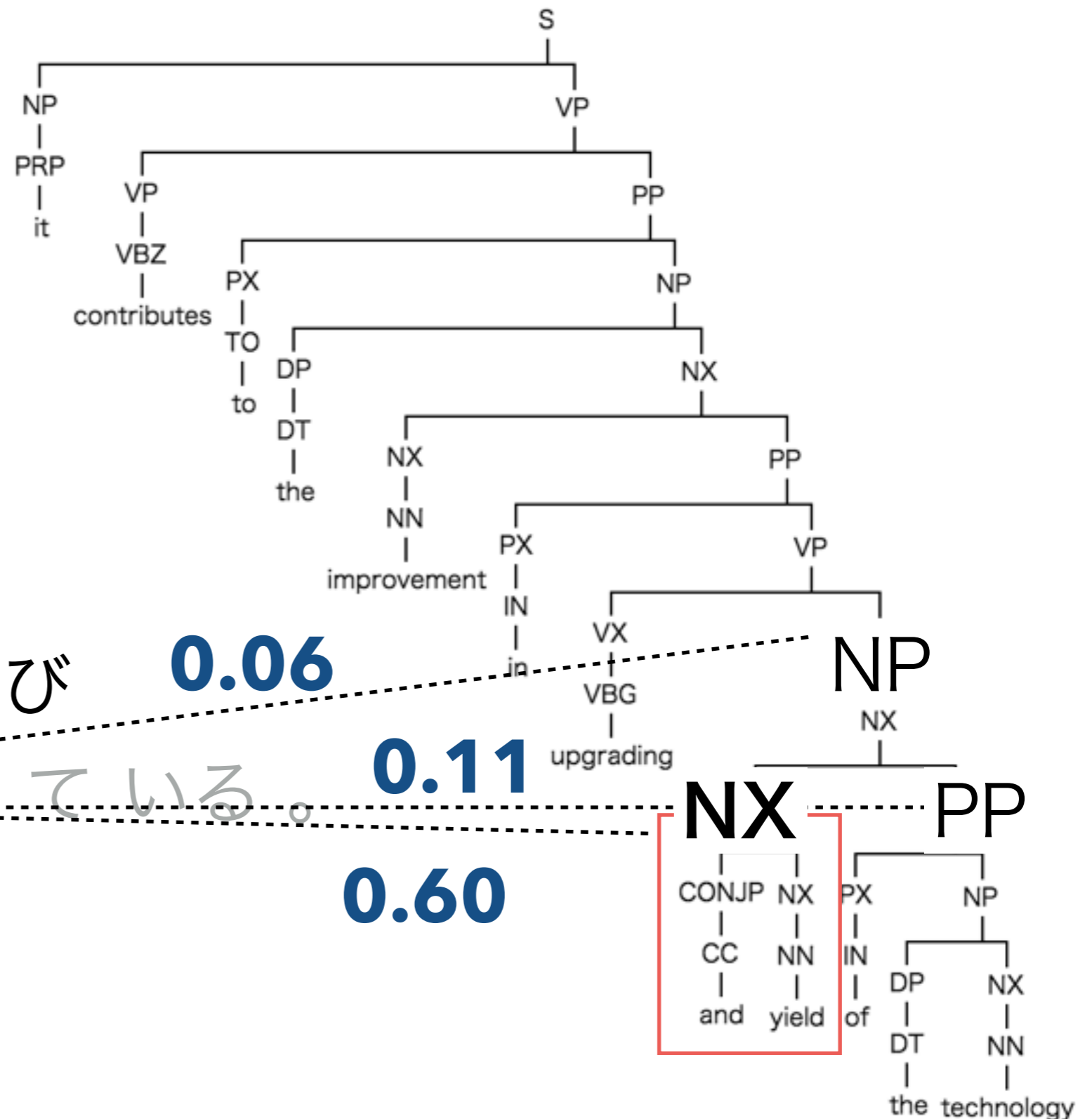
翻訳結果の例2



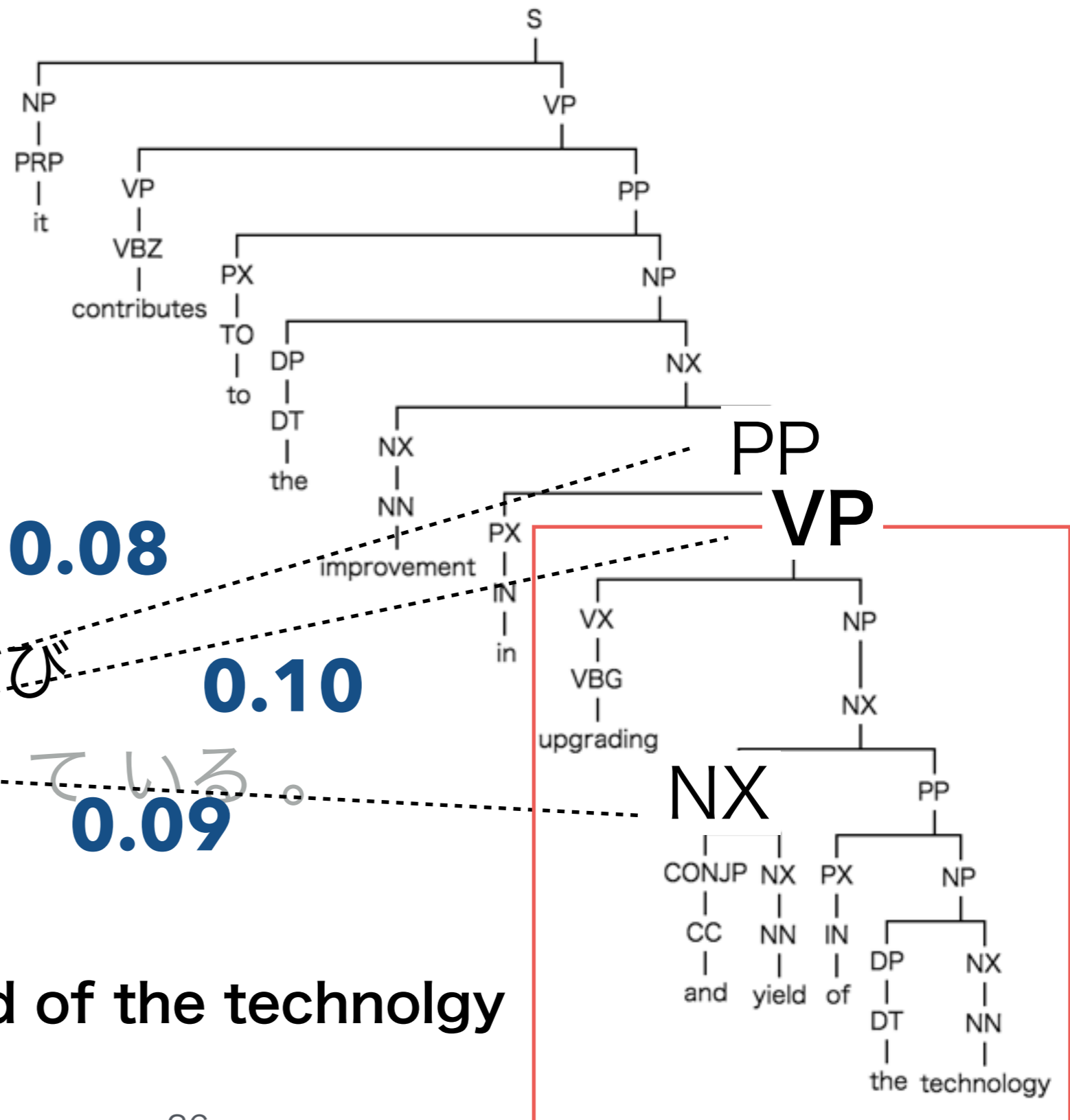
技術の高度化および
収量向上に貢献している。

最注目箇所:
[PP] and

翻訳結果の例2



翻訳結果の例2

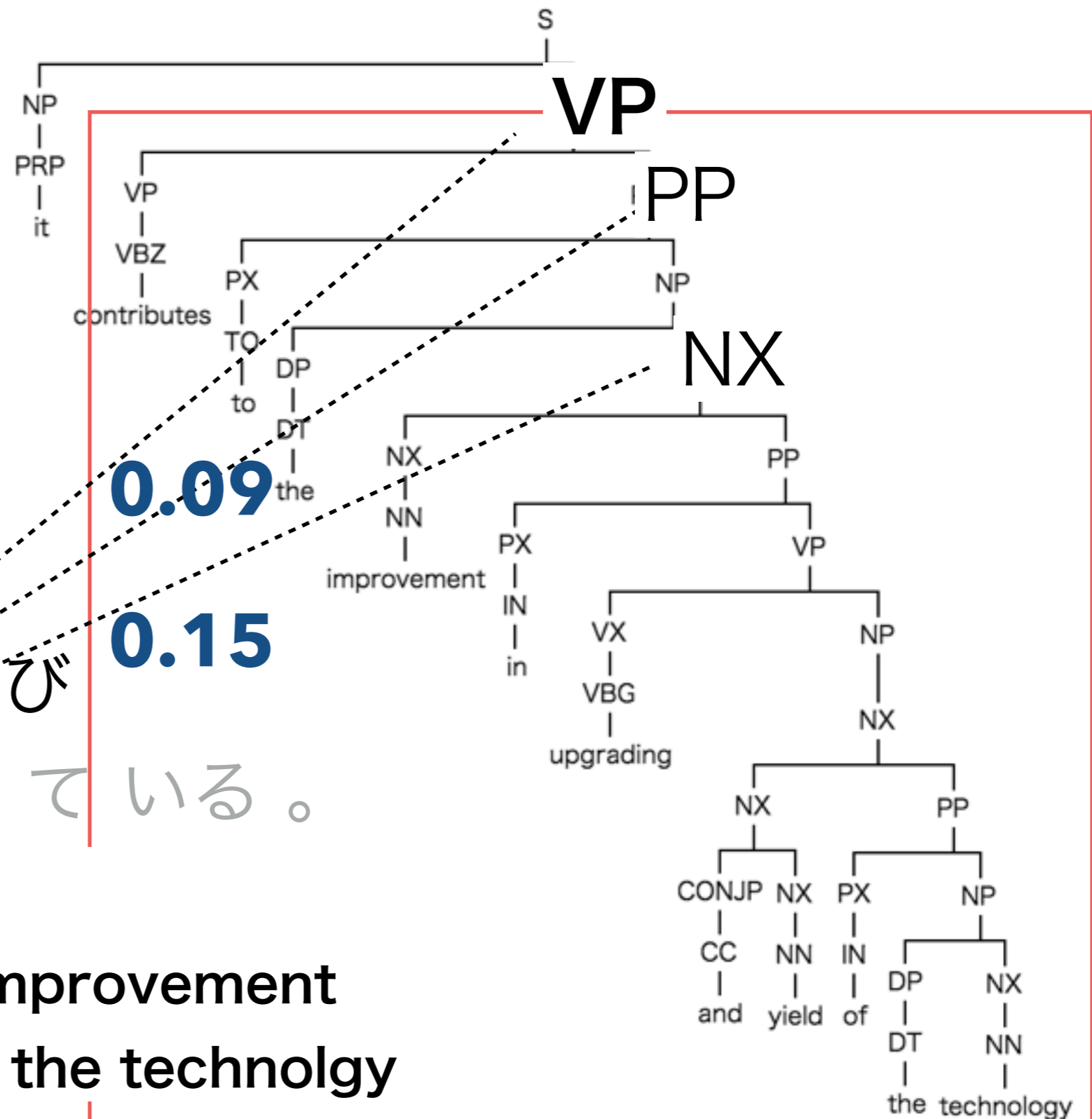


技術の高度化および
収量向上に貢献している。

最注目箇所:

[VP] upgrading and yield of the technology

翻訳結果の例2

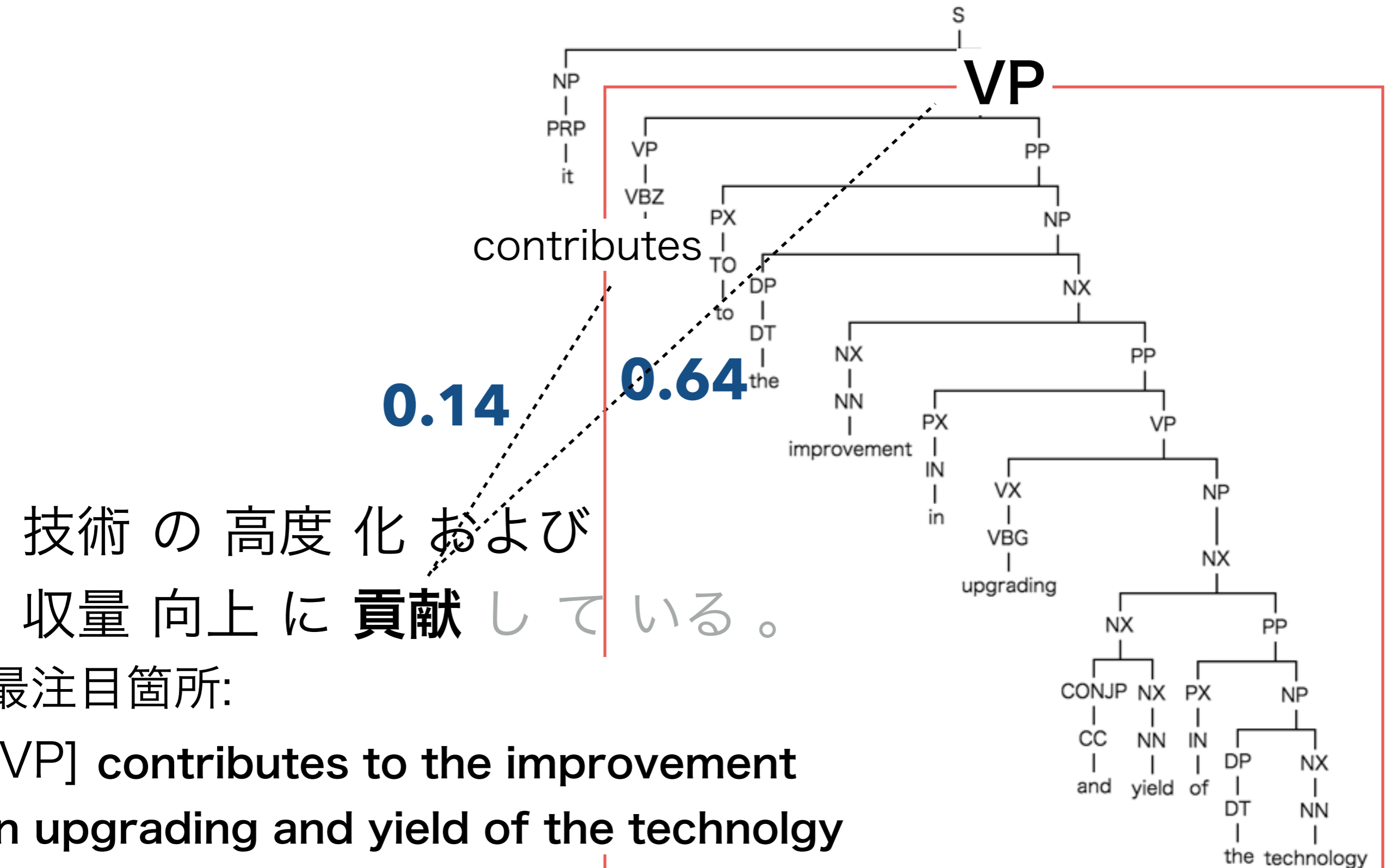


技術の高度化および
収量向上に貢献している。

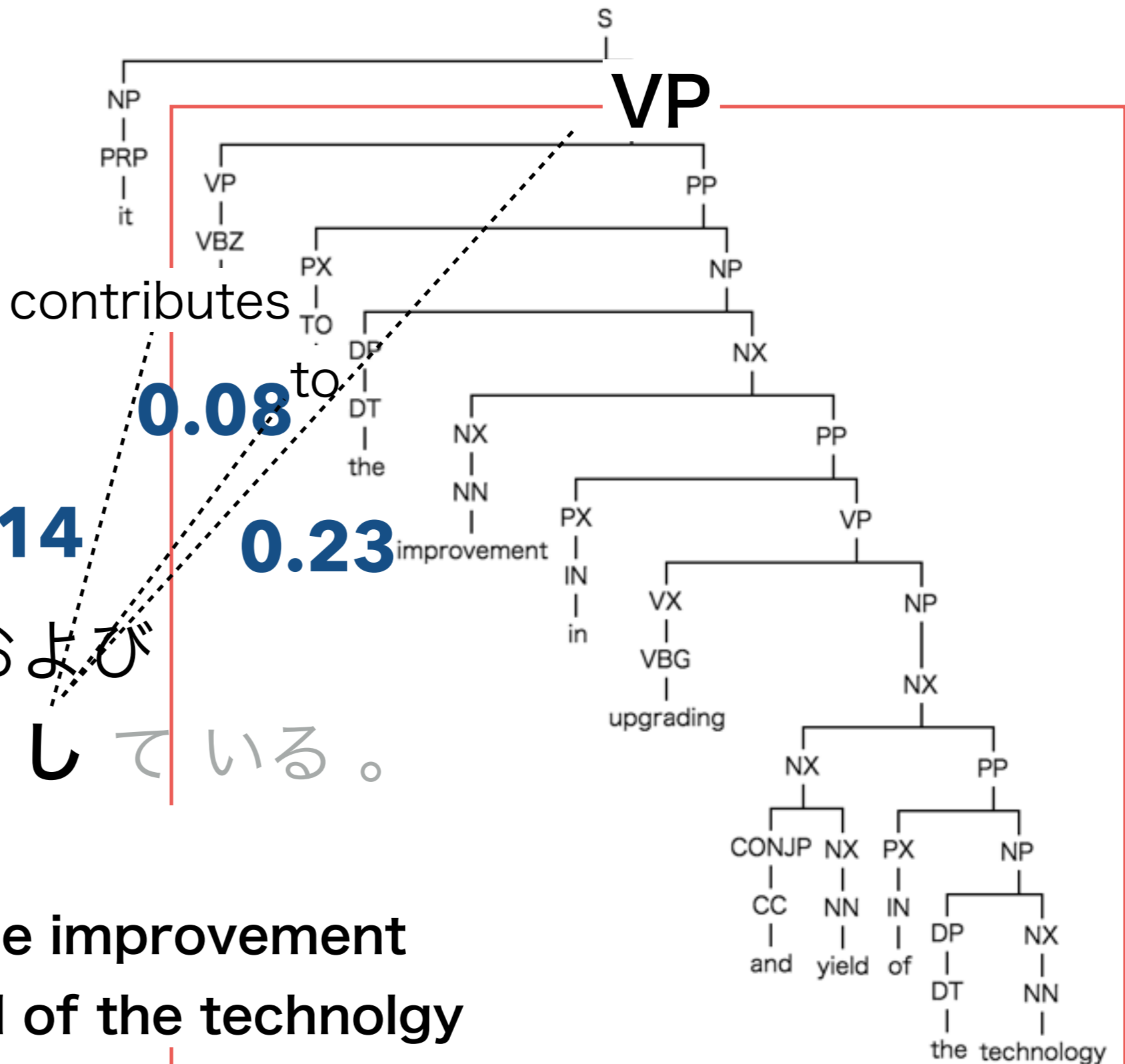
最注目箇所:

[VP] contributes to the improvement
in upgrading and yield of the technology

翻訳結果の例2



翻訳結果の例2

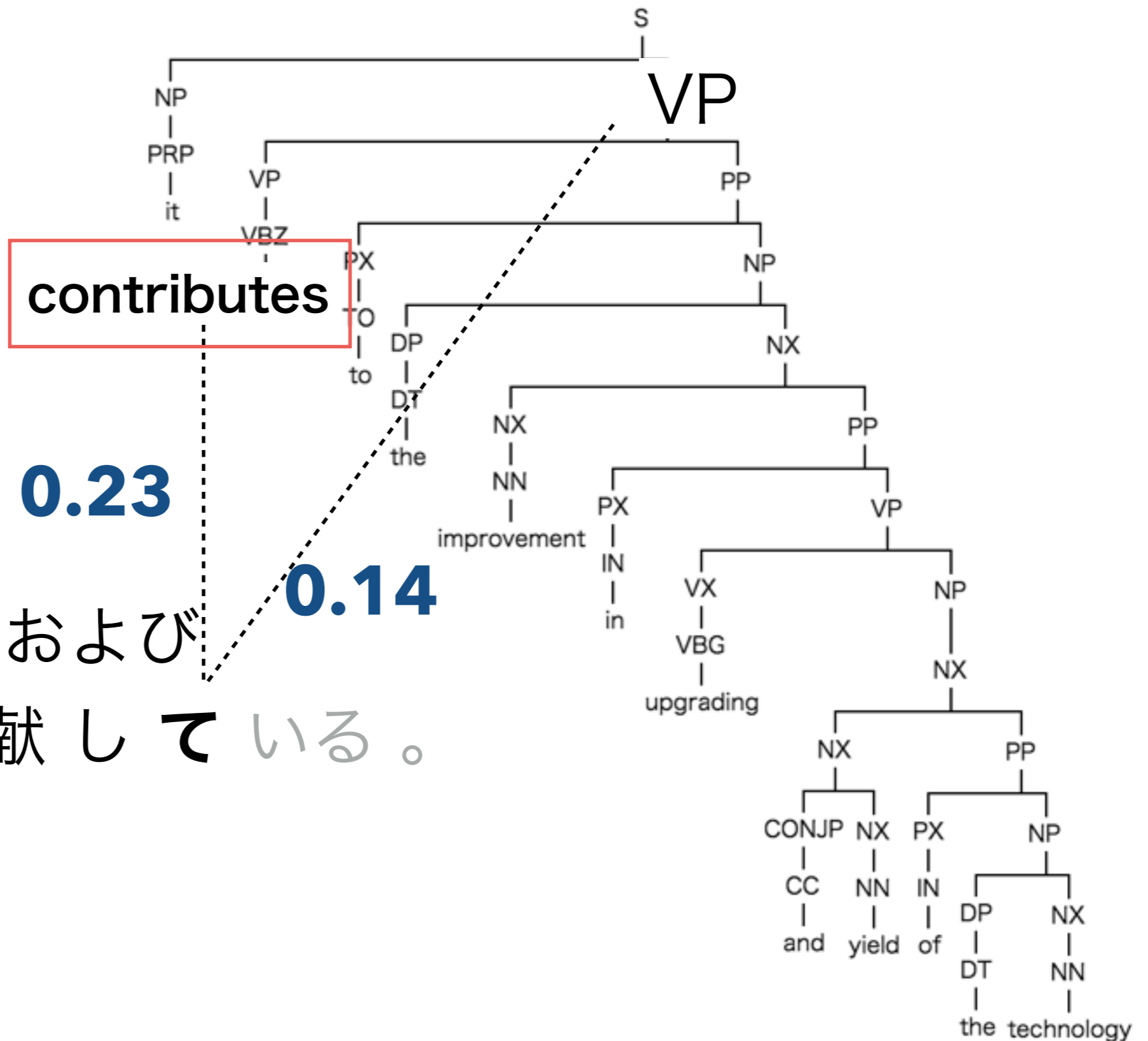


技術の高度化および
収量向上に貢献している。

最注目箇所:

[VP] contributes to the improvement
in upgrading and yield of the technology

翻訳結果の例2

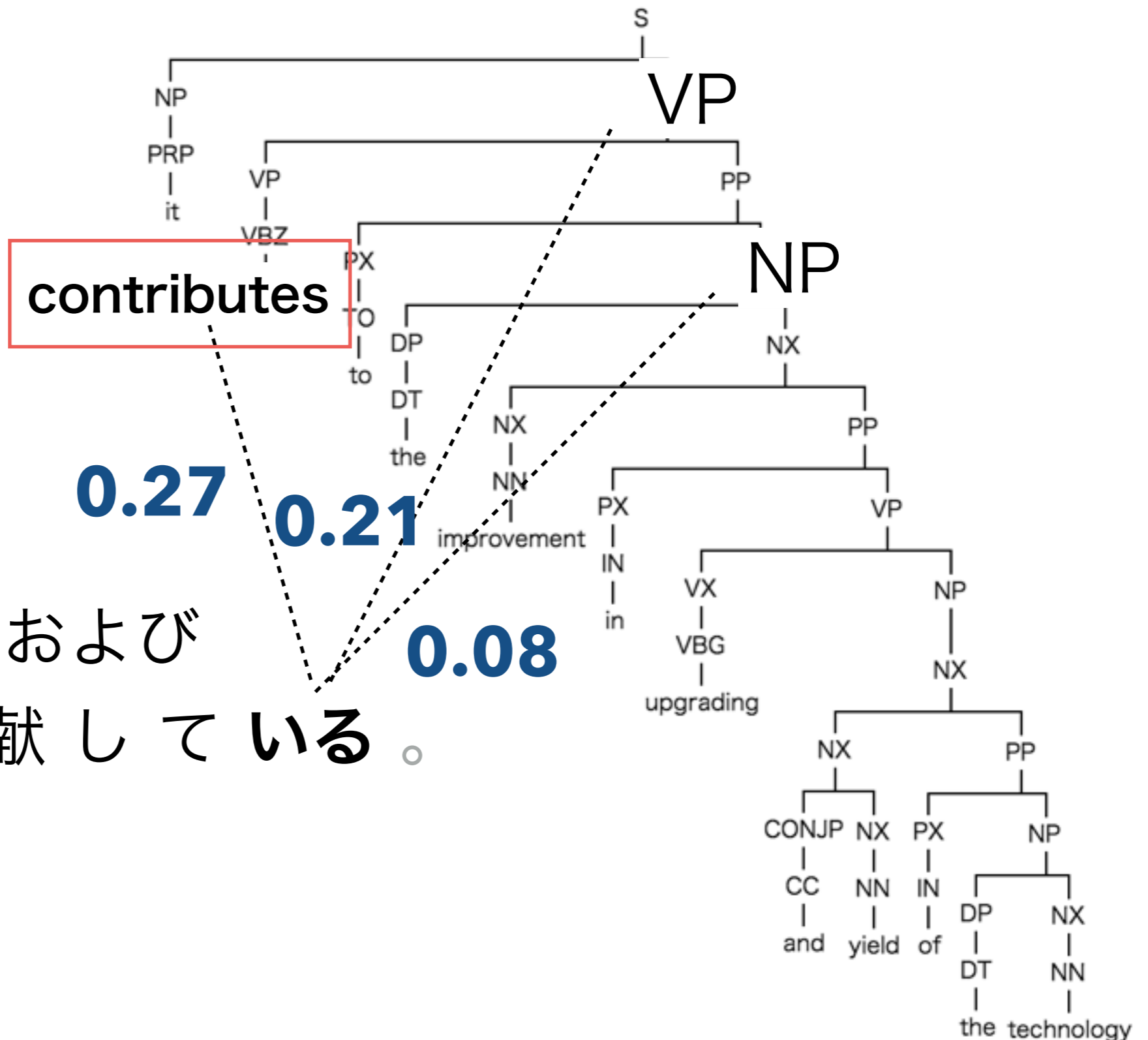


技術の高度化および
収量向上に貢献している。

最注目箇所:

[PP] contributes

翻訳結果の例2

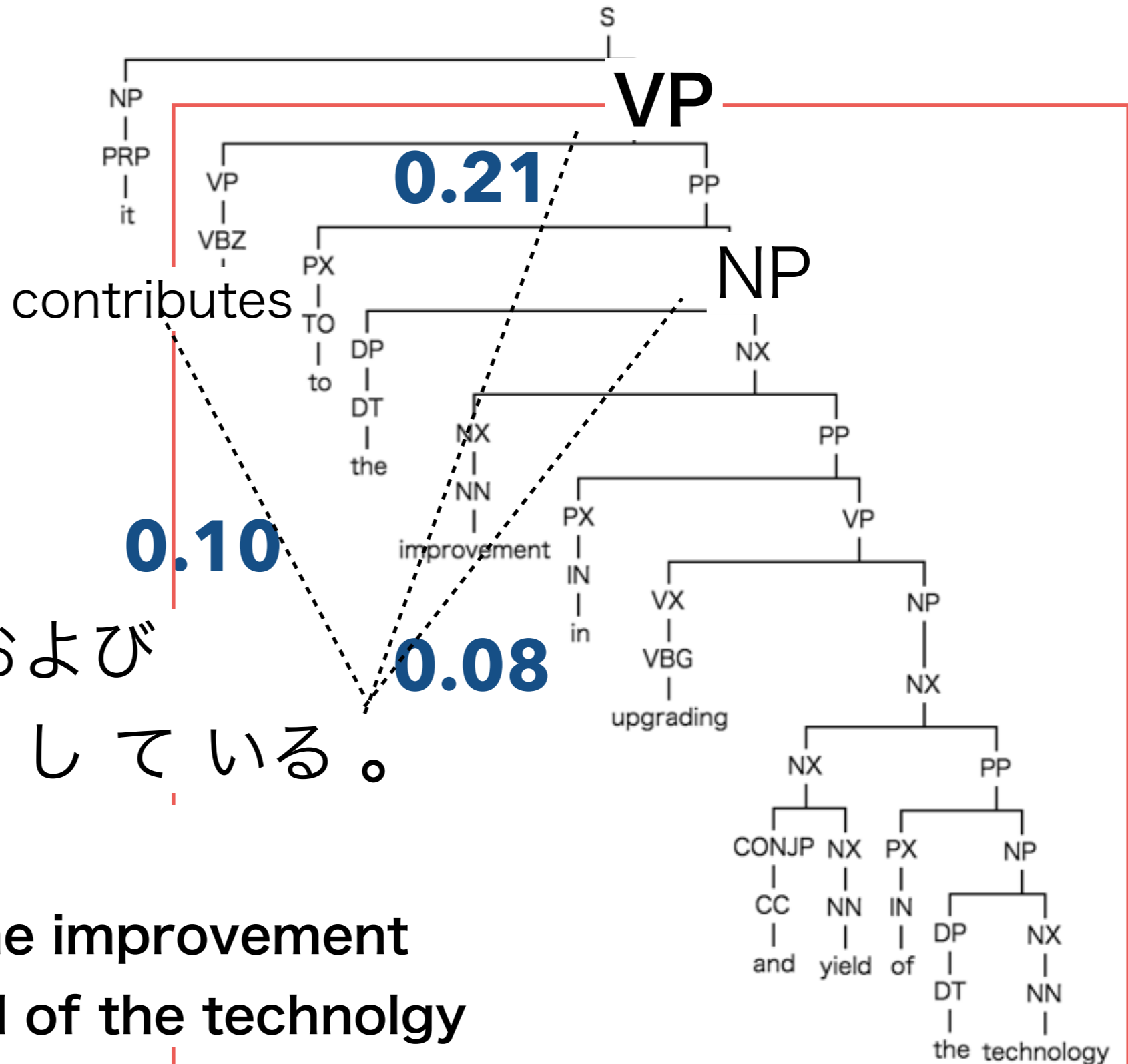


技術の高度化および
収量向上に貢献している。

最注目箇所:

[PP] contributes

翻訳結果の例2

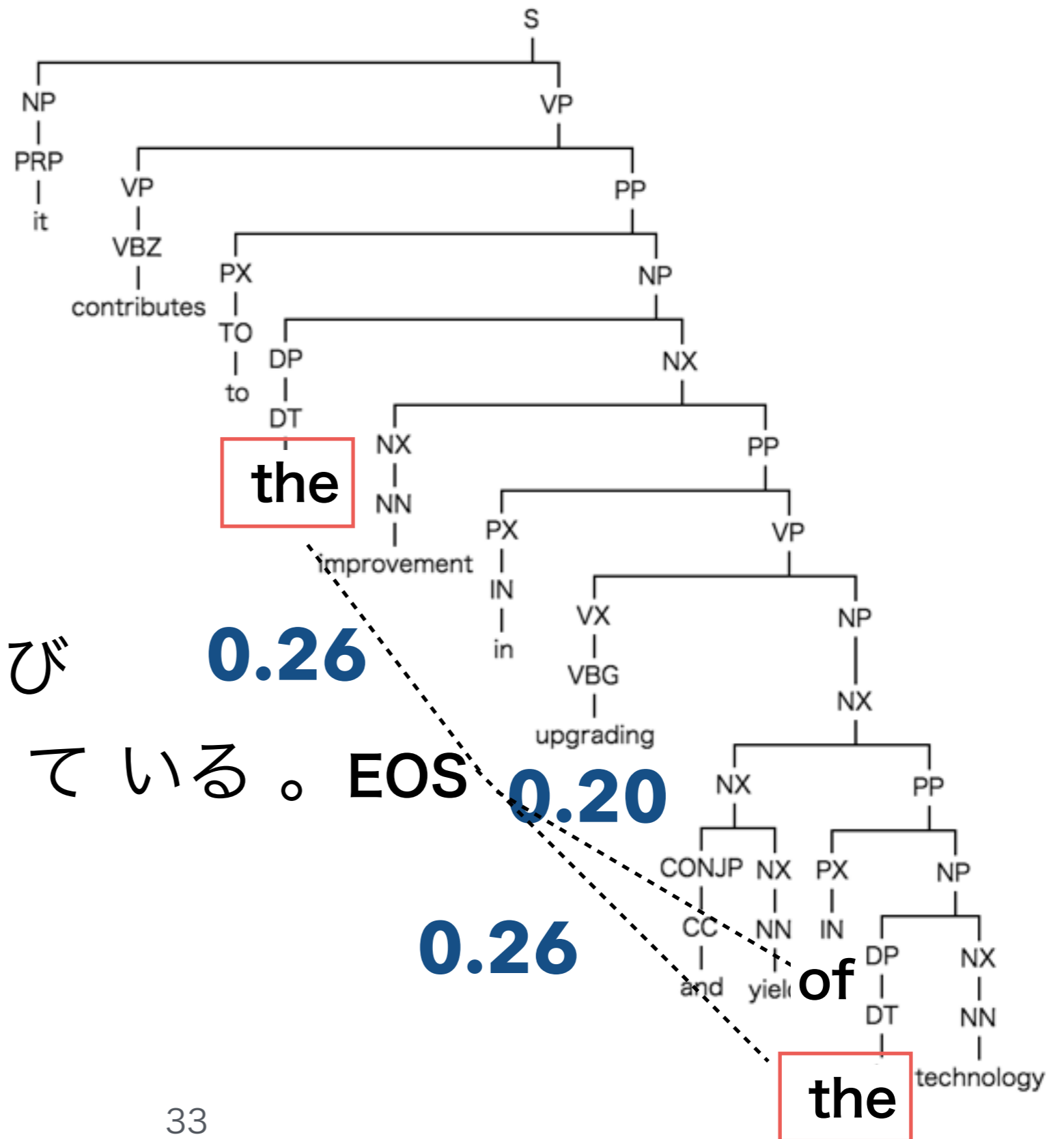


技術の高度化および
収量向上に貢献している。

最注目箇所:

[VP] contributes to the improvement
in upgrading and yield of the technology

翻訳結果の例2



技術の高度化および

収量向上に貢献している。EOS

0.26

0.20

0.26

最注目箇所:

[VP] the

目次

- 背景・目的
- 提案手法
 - 句構造のためのエンコーダ
 - 句構造へのアテンション機構
- 実験・評価
- **まとめと今後の課題**

まとめと今後の課題

- まとめ
 - 句構造へのアテンションに基づく翻訳モデルを提案
 - 英日翻訳コーパスによる実験
 - 句構造のみを利用した提案モデルが最良
 - 原言語の句/ 単語と目的言語の単語の対応を同時学習
- 今後の課題
 - 句構造以外の情報の利用
 - デコーダ側の工夫