

Modeling and Optimizing on Syntactic N-grams for Statistical Machine Translation

**Rico Sennrich
TACL 2015**

江里口瑛子

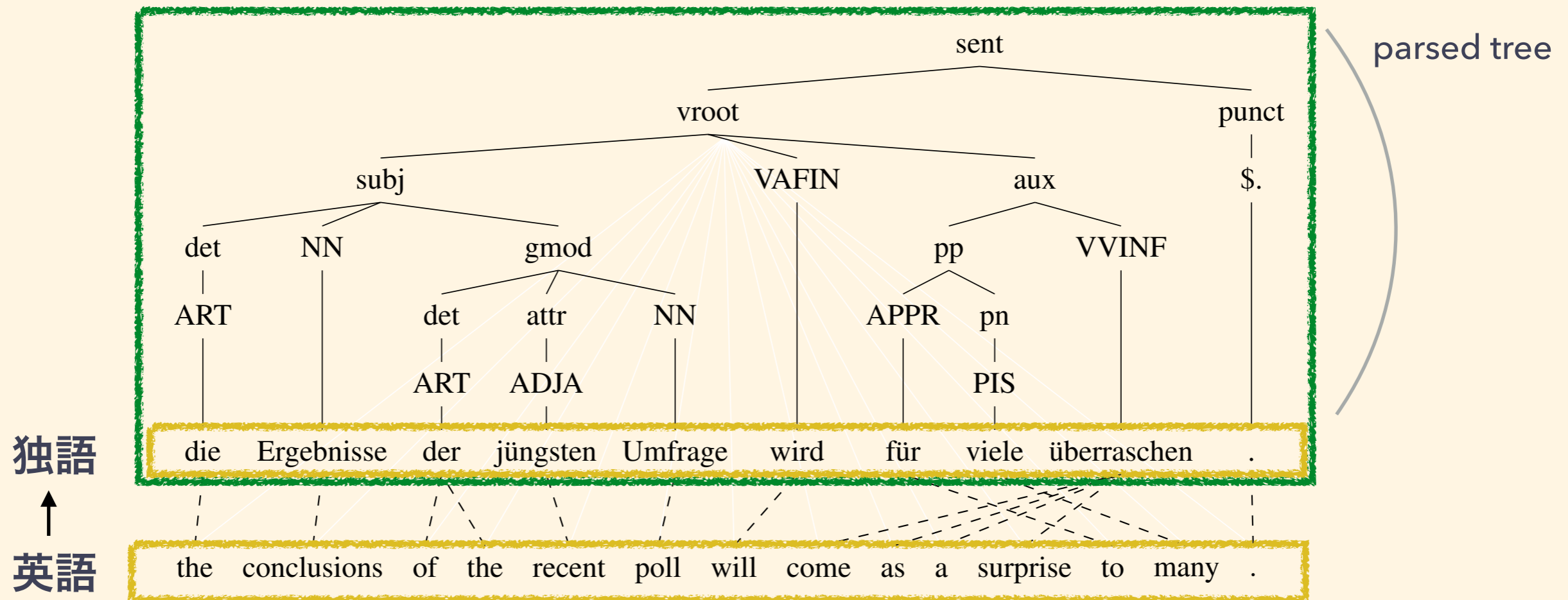
東京大学 鶴岡研究室 博士課程1年

概要

- 従来のn-gram言語モデルベースでは、流暢な翻訳表現は獲得できていない
 - 翻訳評価指標のBLEU (n-gram-base)
 - cf. syntactic n-gram [Sidorov et al., 2013]
- 単語の係り受け情報を利用した言語モデルの学習
 - Dependencyモデル [Shen et al., 2010]
 - Neural言語モデル [Bengio et al., 2003]
 - string-to-tree翻訳モデルで翻訳結果を出力

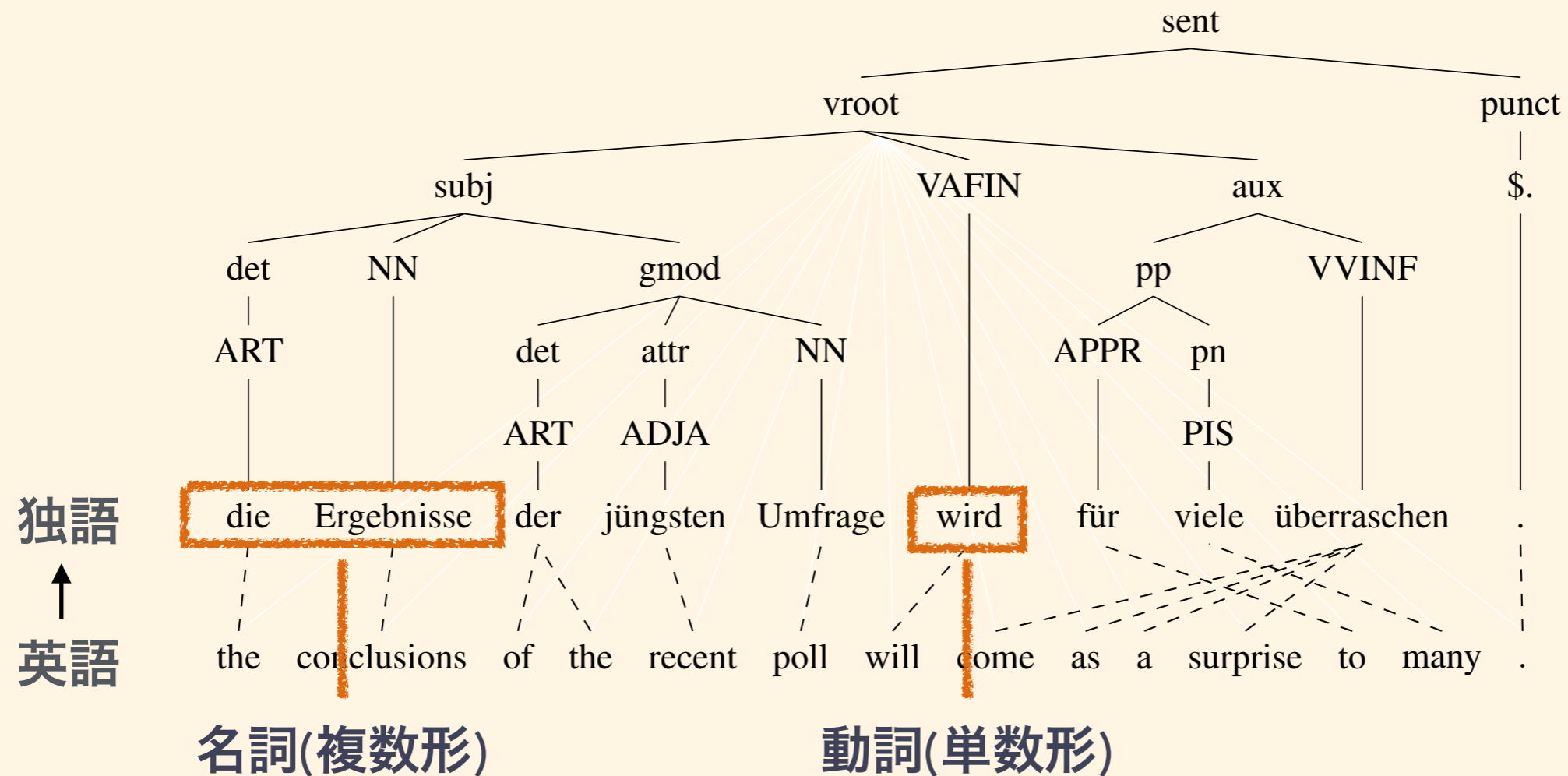
string-to-tree 翻訳結果

- 出力先がtreeなので翻訳結果の構文が安定



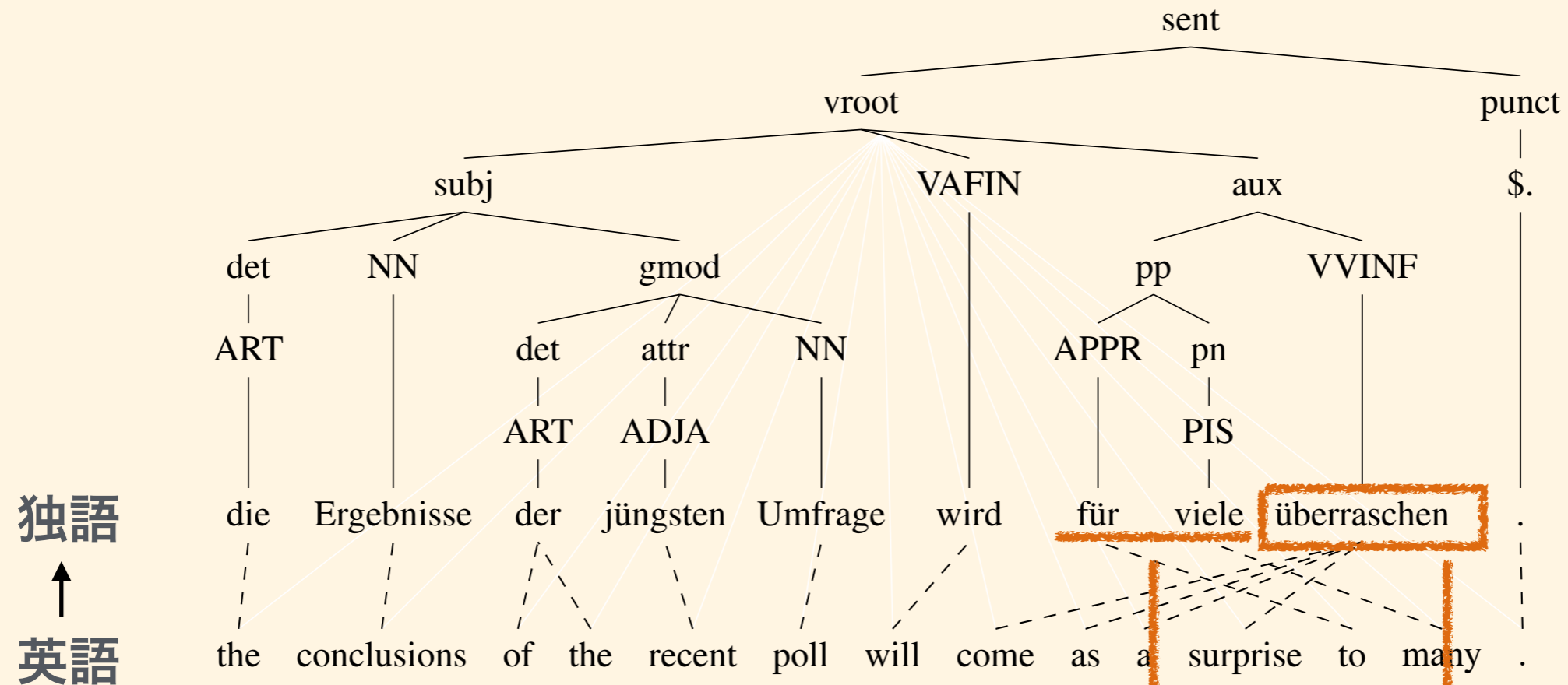
string-to-tree 翻訳結果の課題1

- Morphological agreement error



string-to-tree 翻訳結果の課題2

- Subcategorization error

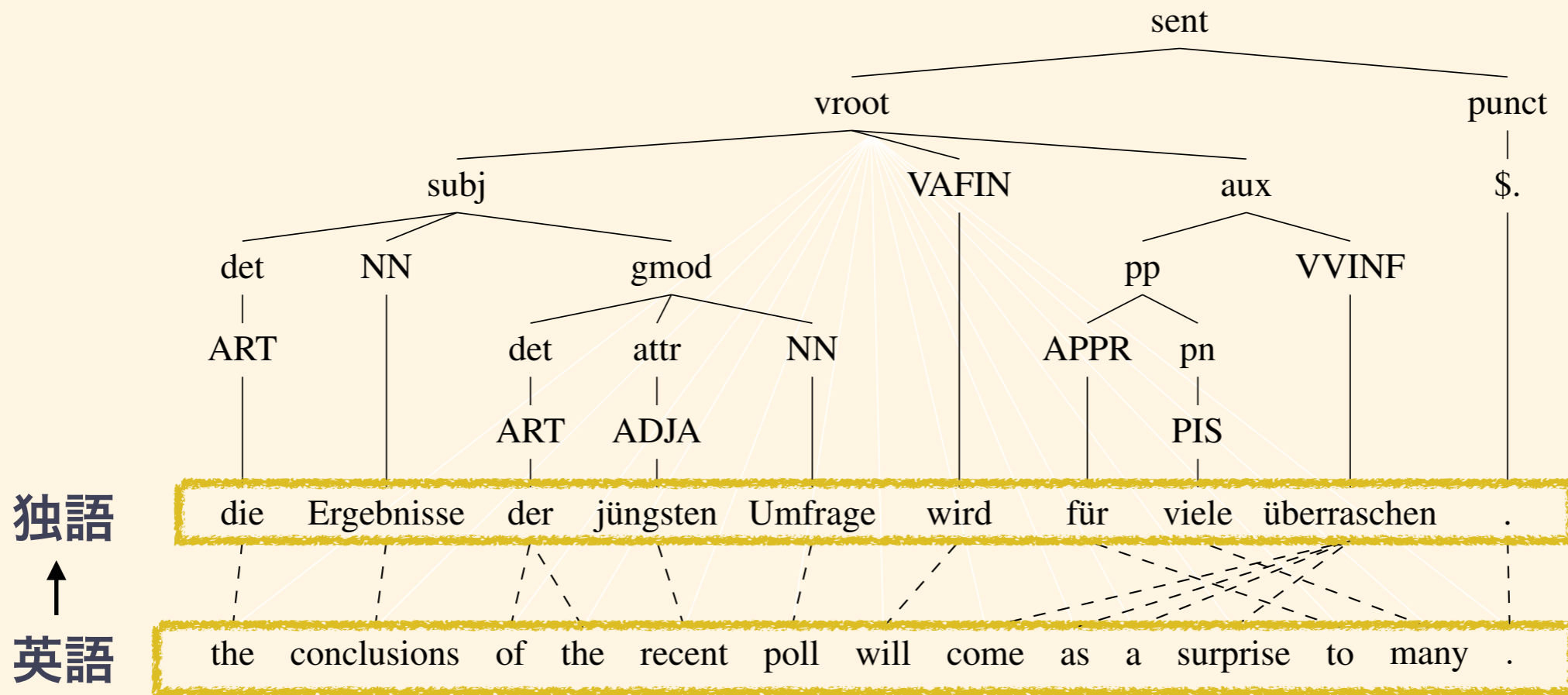


他動詞(überraschen)の 他動詞

目的語がない

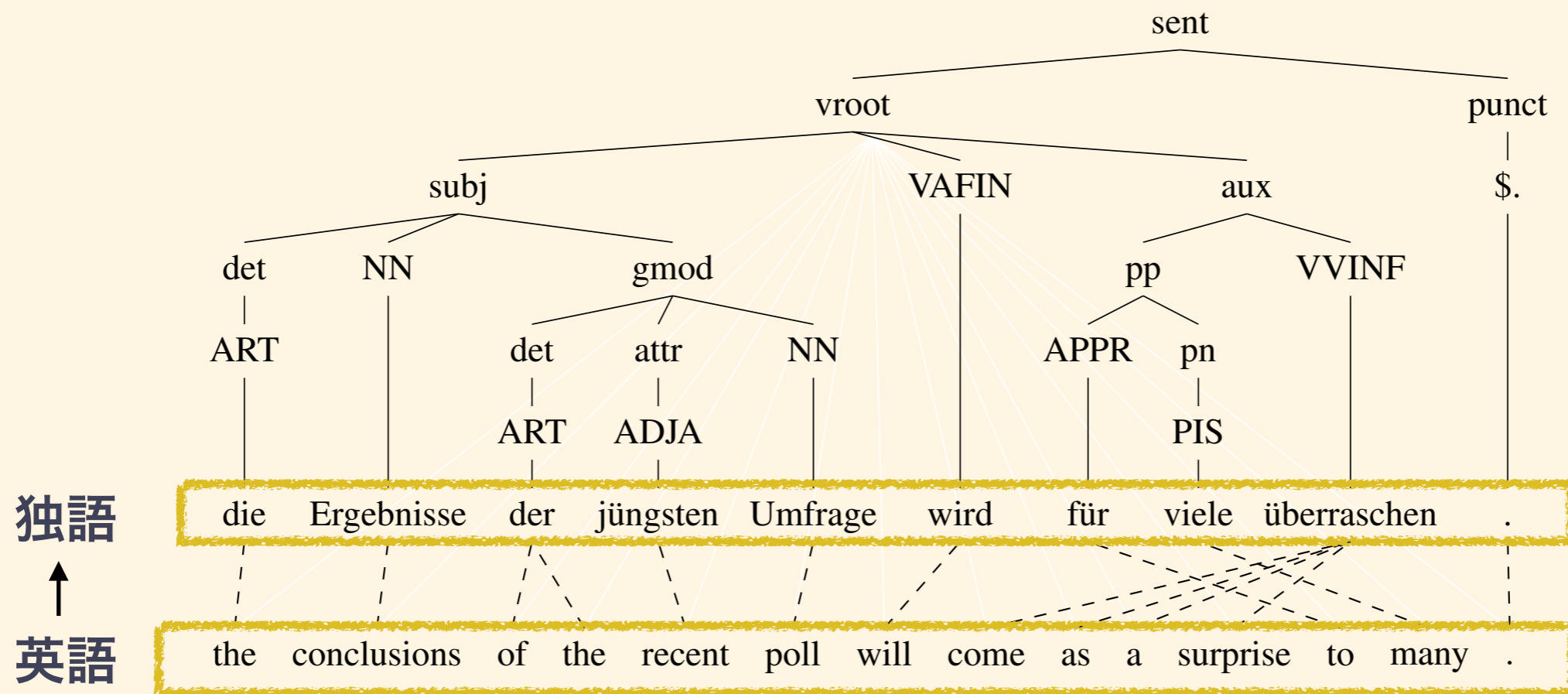
string-to-tree 翻訳結果の課題

- (relatively) free word orderな言語 (独, 露..)
- 係り受け順+係り受けラベル



string-to-tree 翻訳結果の課題

- (relatively) free word orderな言語 (独, 露..)
- 係り受け順+係り受けラベル



Dependency Language Model; DLM

- $S = \{w_1, w_2, \dots, w_n\}$

$$P(S) \approx \prod_{i=1}^n P(w_i | h_s(i)_1^q, h_a(i)_1^r)$$

Dependency Language Model; DLM

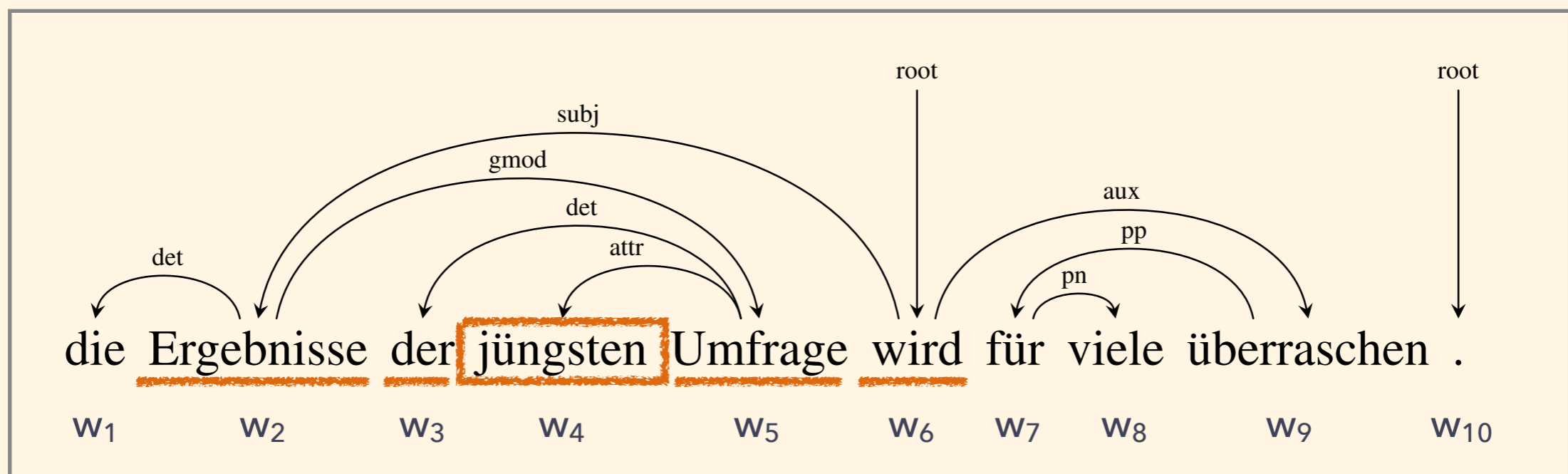
- $S = \{w_1, w_2, \dots, w_n\}$

$w_4 = \text{jüngsten}$

$h_{\text{sibling}}(4) = (\text{der})$

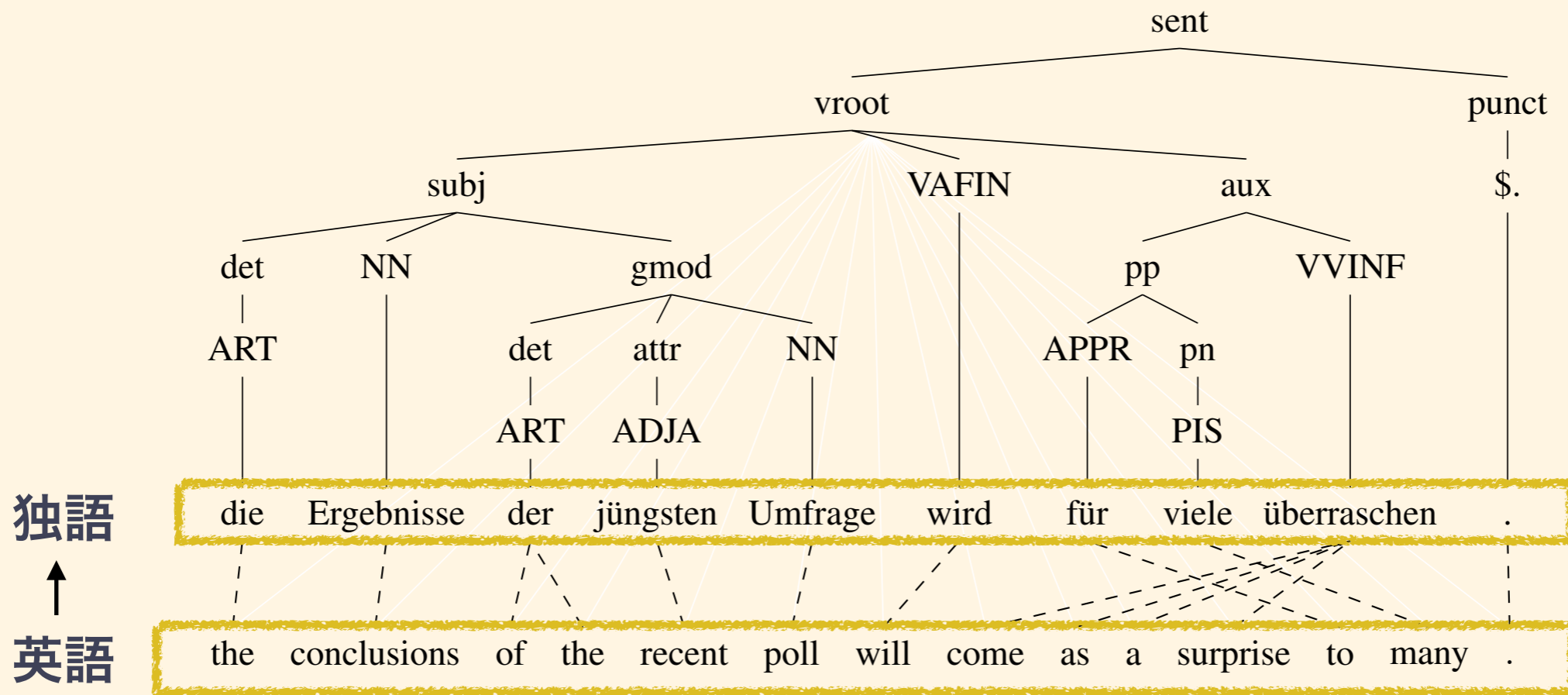
$h_{\text{ancestor}}(4) = (\text{Umfrage, Ergebnisse, wird, } \varepsilon)$

$$P(S) \approx \prod_{i=1}^n P(w_i | h_s(i)_1^q, h_a(i)_1^r)$$



string-to-tree 翻訳結果の課題

- (relatively) free word orderな言語 (独, 露..)
- 係り受け順+係り受けラベル



Relational Dependency Language Model; RDLM

- $D = \{l_1, l_2, \dots, l_n\}$

$$P(S, D) = P(D) \times P(S|D)$$

$$\approx \prod_{i=1}^n P_l(i) \times P_w(i)$$

係り受け単語列
係り受けラベル列

$$P_l(i) = P(l_i | h_s(i)_1^q, l_s(i)_1^q, h_a(i)_1^r, l_a(i)_1^r)$$

$$P_w(i) = P(w_i | h_s(i)_1^q, l_s(i)_1^q, h_a(i)_1^r, l_a(i)_1^r, l_i)$$

Relational Dependency Language Model; RDLM

- $D = \{l_1, l_2, \dots, l_n\}$

$$P(S, D) = P(D) \times P(S|D)$$

$$\approx \prod_{i=1}^n P_l(i) \times P_w(i)$$

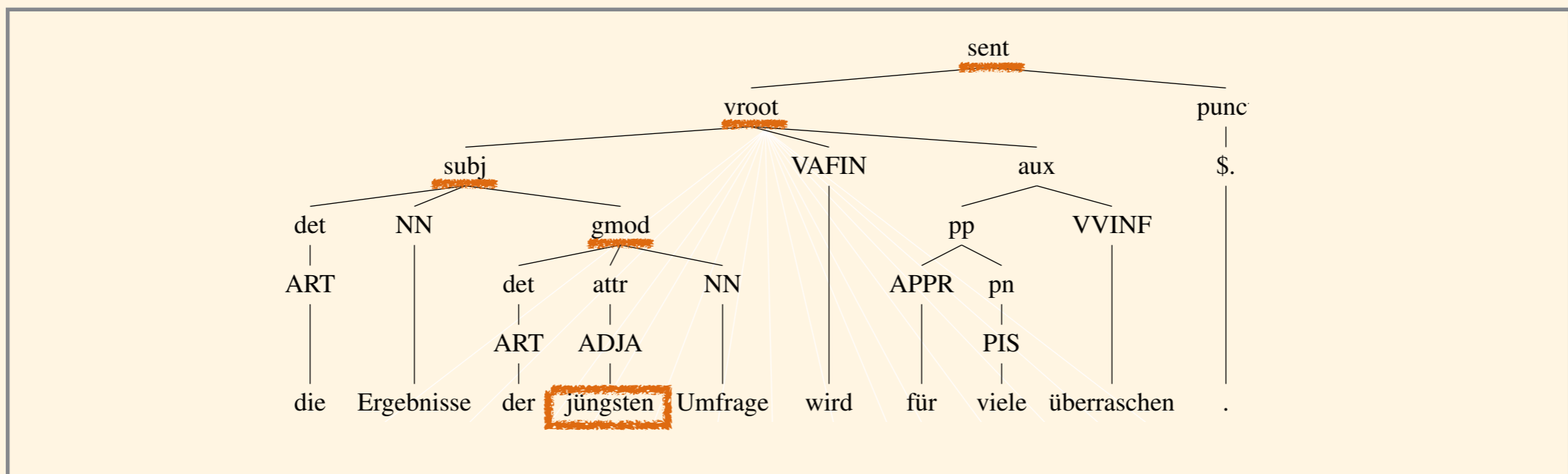
$$P_l(i) = P(l_i | h_s(i)_1^q, l_s(i)_1^q, h_a(i)_1^r, l_a(i)_1^r)$$

$$P_w(i) = P(w_i | h_s(i)_1^q, l_s(i)_1^q, h_a(i)_1^r, l_a(i)_1^r, l_i)$$

$l_4 = \text{attr}$

$l_{\text{sibling}}(4) = (\text{det})$

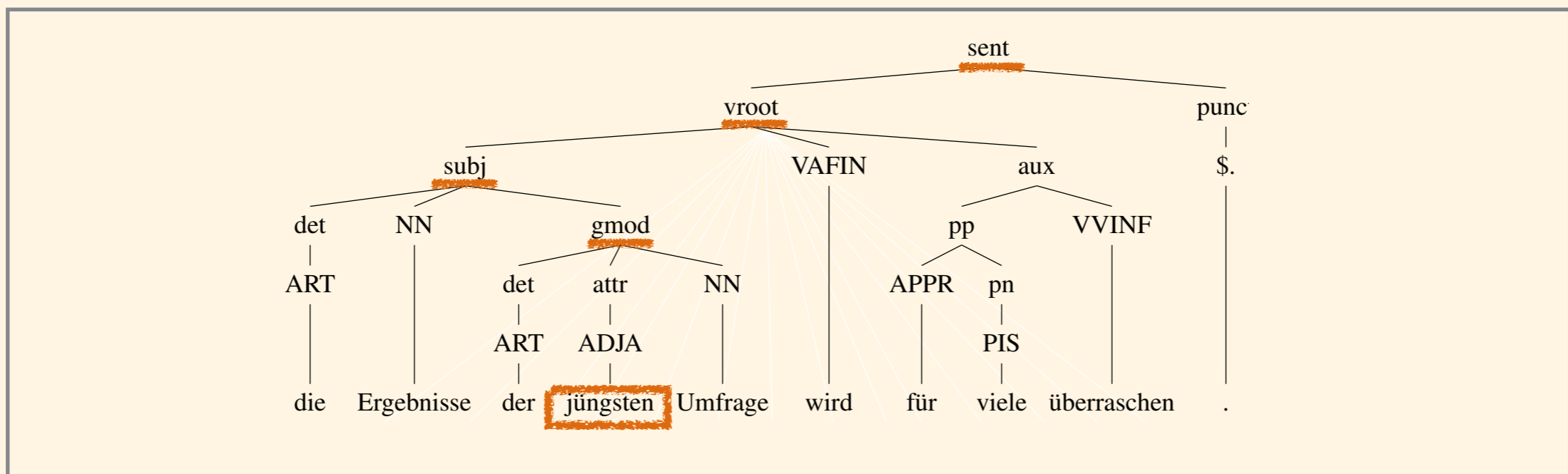
$l_{\text{ancestor}}(4) = (\text{gmod}, \text{subj}, \text{vroot}, \text{sent})$



Relational Dependency Language Model; RDLM

- Tree topology: T

$$P(S, D, T) \approx \prod_{i=1}^m \begin{cases} P_l(i) \times P_w(i), & \text{if } w_i \neq \epsilon \\ P_l(i), & \text{otherwise} \end{cases} \quad (4)$$



Neural Network Training

- Standard feed-forward neural network [Bengio et al., 2003]
 - 隠れ層1, 活性化関数(rectified-linear)
 - 単語 P_w とラベル P_l をそれぞれ学習
- string-to-treeモデルに言語モデルの素性として追加

実験

- SMT 2014 shared Task (英独, 英露)
- string-to-treeシステム with Moses
- MIRA object:
 - BLEU (n-gram; N=4)
 - HWCM_f (syntactic n-gram; N=4)
- 評価指標:
 - BLEU, HWCM, METEOR, TER

各言語モデルのPerplexity

- 参照訳(ref)とbaselineシステム(1-bestの出力)に対するパープレキシティ

language model	perplexity		entropy difference
	ref.	1-best	
5-gram (KN)	232.9	183.3	-4.4%
5-gram NNLM	207.3	207.5	0.0%
Shen et al. (2010)	345.1	383.0	1.8%
DLM ($q=1; r=1$)	213.7	259.9	3.6%
DLM ($q=1; r=2$)	136.9	188.3	6.5%
RDLM ($q=1; r=2$)	349.2	734.6	12.7%
RDLM, P_w	58.1	85.1	9.4%
RDLM, P_l	6.0	8.6	20.1%

実験結果

- 英語→ドイツ語

- BLEU, HWCM, MENTOR (↑), TER (↓)

MIRA objective	system	dev				newstest2013				newstest2014			
		BLEU	HWCM _f	METEOR	TER	BLEU	HWCM _f	METEOR	TER	BLEU	HWCM _f	METEOR	TER
BLEU	baseline	34.4	32.6	52.5	47.4	19.8	22.8	39.7*	62.4	20.3	23.2	42.0*	62.7
	5-gram NNLM	35.3	33.1	53.2*	46.4	20.4	23.2	40.2	61.7	21.0	23.5	42.5*	62.2
	Shen et al. (2010)	34.4*	33.2	52.7*	46.9	20.0	23.2	40.0*	62.3	20.4	23.5	42.3*	62.9
	DLM	34.9*	33.8	53.1*	46.8	20.3	23.6	40.1*	61.7	20.8	23.9	42.3*	62.2
	RDLM	35.0	33.9	53.1*	46.7	20.5	23.8	40.4*	61.7	21.0	24.1	42.7*	62.2
	5-gram + RDLM	35.5	34.0	53.4*	46.3	20.7	23.7	40.6*	61.5	21.4	24.1	42.9*	61.7
BLEU + HWCM _f	baseline	34.4	33.0*	52.4	46.9*	20.0*	23.0*	39.6	61.9*	20.5*	23.3*	41.8	62.2*
	5-gram NNLM	35.2	33.5*	53.0	46.0*	20.6*	23.4*	40.1	60.9*	21.1*	23.6	42.3	61.5*
	Shen et al. (2010)	34.2	33.8*	52.4	46.4*	20.2*	23.5*	39.8	61.8*	20.7*	23.7*	42.1	62.2*
	DLM	34.8	34.3*	52.7	45.9*	20.4	23.8*	39.8	60.7*	21.4*	24.2*	42.0	60.9*
	RDLM	34.9	34.5*	53.0	45.8*	20.9*	24.2*	40.3	60.7*	21.6*	24.5*	42.5	60.8*
	5-gram + RDLM	35.4	34.6*	53.2	45.4*	21.0*	24.1*	40.4	60.5*	21.8*	24.4*	42.7	60.6*

翻訳結果の例

1	source baseline best reference	also the user manages his identity and can therefore be anonymous. auch der Benutzer verwaltet seine Identität und können daher anonym sein. auch der Benutzer verwaltet seine Identität und kann daher anonym sein. darüber hinaus verwaltet der Inhaber seine Identität und kann somit anonym bleiben.
2	source baseline best reference	how do you apply this definition to their daily life and social networks? wie kann man diese Definition für ihr tägliches Leben und soziale Netzwerke gelten ? wie kann man diese Definition auf ihren Alltag und sozialen Netzwerken anwenden ? wie wird diese Definition auf seinen Alltag und die sozialen Netzwerke angewendet ?
3	source baseline best reference	the City Council must reach a decision on this in December. Der Stadtrat muss im Dezember eine Entscheidung darüber erzielen . Im Dezember muss der Stadtrat eine Entscheidung darüber treffen . Im Dezember muss dann noch die Stadtverordnetenversammlung entscheiden .

- 1. subject-verb error (単数形-単数形)
- 2. subcategorization error (anwenden: 他動詞)
- 3. collocation error (適切な語彙選択)

Morphological Agreement

- newstest2013で3000文中で、morphological agreement errorが1つでも含まれた文の数

system	MIRA objective	
	BLEU	BLEU+HWCM _f
baseline	1028	1018
5-gram NNLM	845	825
Shen et al. (2010)	884	844
DLM	680	599
RDLM	550	468
5-gram + RDLM	576	484

まとめ

- 単語の係り受け情報を利用した言語モデルの学習
 - syntactic n-gram言語モデル (RDLM)の提案
 - string-to-tree翻訳への適用 (英→独; 英→露)
 - 翻訳出力結果の改善
 - dependency言語モデル評価はHWCMスコア
- Moses decoderの一部として実装公開
(<http://www.statmt.org/moses/?n=FactoredTraining.BuildingLanguageModel#ntoc43>)