

# What Determines the Order of Adjectives in English? Comparing Efficiency-Based Theories Using Dependency Treebanks

**Richard Futrell**

University of California, Irvine

`rfutrell@uci.edu`

**William Dyer**

Oracle Corporation

`william.dyer@oracle.com`

**Gregory Scontras**

University of California, Irvine

`g.scontras@uci.edu`

最先端NLP 2020

発表者: 神藤駿介 (東大 宮尾研 M1)

# 自己紹介

- 宮尾研究室所属 M1 (去年の9月から)
- 卒論: NNモデルの獲得する言語知識の検証
  - COLING 2020 submitted
- 今は言語の普遍的な性質に漠然と興味がある
  - もともと応用よりは言語そのものに興味あり
  - 何が言語を言語たらしめているのか？
- 「理学的」な研究がしてみたい
  - 謎とされている存在の輪郭を描き出したい( NN然り、数多の言語現象然り)

# 概要

▶ テーマ: 英語のテキストの形容詞の順序を決める要因を調査

▶ 形容詞には認知的に「好ましい」順序がある

⇒ 順序を決めるなんらかの制約・理論があるはず ...

😊 "big blue box"

V.S.

😞 "blue big box"

## Efficiency-based theory

仮定: 言語は人が「効率的に」処理できるようにできているはず!

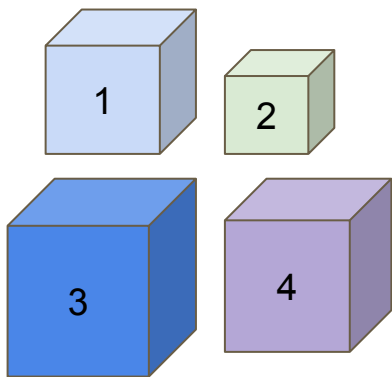
# Efficiency-based Theory

- ▶ 仮定：言語は人が「効率的に」処理できるようにできているはず
  - ⇒ 形容詞の順序は人の認知負荷を小さくするような制約によって決まる
    - Subjectivity
    - Information locality
    - Integration cost
    - Information gain

これら4つの理論について定量的に評価・比較

# Subjectivity

- ▶ 主観的な形容詞ほど前(名詞から離れた位置)に来る (Hatzron, 1978)
  - "blue" より "big" の方が主観的なので "big blue box" が好ましい
    - 「大きさ」の基準は人に寄りそうけど、「青さ」の基準は人によってそれほど差がなさそう
  - 利点: コミュニケーションの齟齬が起きづらくなる (Franke et al., 2019)



例) 3 について言及したい場合...

- "big blue box" と発話 : [big [blue box]]
  - 聞き手「青い箱は 1 か 3 だ。そのうち大きいのは 3 だ」
- "blue big box" と発話 : [blue [big box]]
  - 聞き手「大きい箱は 2 以外だな。1 と 3 のどっちのことだろう？」

# Information locality

- ▶ 名詞との PMI が大きい形容詞ほど後ろ(名詞の近くに)来る
  - 情報をより効率的に伝えたい ⇨ surprisal を小さくしたい
  - よくある形容詞-名詞のペアが近くなるように配置すれば良い

$$\text{PMI}(a : n) \equiv \log \frac{p(a, n)}{p(a)p(n)}.$$

- $p(a, n)$  ... 依存関係にある adj, noun の分布
- $p(a), p(n)$  ...  $p(a, n)$  の周辺確率

# Integration cost

- ▶ 修飾可能な名詞の数が少ない形容詞ほど後ろに来る (Ziff, 1960)

例) "big wooden spoon"

- big ... 修飾できる名詞はいろいろ思いつく
- wooden ... 修飾できる名詞は限られそう
- 修飾可能な名詞が多い = 曖昧性が大きい ⇨ エントロピーで表現

$$\text{Cost}(w) \propto H[T|w]$$

$w$  ... 形容詞

$$= \sum_{t \in T} -p_T(t|w) \log p_T(t|w),$$

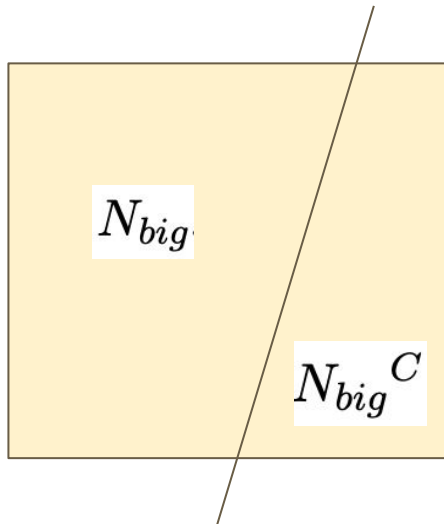
$T$  ... 修飾できる名詞の集合

- エントロピーが単調減少するように形容詞を配置

# Information gain

... 筆者らが提案、Integration cost と似たアイデア

- ▶ ある形容詞によって表現可能な指示対象が単調減少するように配置



(例)"big" によって表現可能な指示対象を定量化したい

- 名詞全体の集合を、"big" で修飾できるもの/できないものに分ける
- 「"big" によって表現可能な指示対象がどれだけ減ったか」を Information gain で定義:

$$\text{IG}(a) = H[N] - \left[ \frac{|N_a|}{|N|} H[N_a] + \frac{|N_a^c|}{|N|} H[N_a^c] \right]$$



# ここまでのまとめ

形容詞の順序を推定する4つの predictor

Predictor	判断基準の要約
Subjectivity	形容詞の主観性の強さ
Information locality	形容詞と名詞のPMI
Integration cost	形容詞が修飾可能な名詞の数
Information gain	表現可能な指示対象の数

⇐ 形容詞単独で判断可能

形容詞と名詞の  
統計情報が必要

# Methods

▶ 目標 ... データから順序の predictor を得たい！

- Scoring function:  $S(A, N)$  を構築すれば良い

- 形容詞(と名詞)を受け取ってスコアを計算、スコアの大小によって順序を定める

1. Subjectivity ... クラウドソーシングで形容詞の主観度を計測

2. それ以外 ... コーパスを用いて計算式に従って得る

- Common Crawl (dependency 付) から以下を得る

- Adjective-Noun (AN) pairs

A  $\xleftarrow{\text{amod}}$  N

- Adjective-Adjective-Noun (ANN) triples

A1      A2  $\xleftarrow{\text{amod}}$  N  
           $\xleftarrow{\text{amod}}$

- Test set ... English web treebank + Common Crawl -> 40K ANN triples

# Predictor の推定方法

## ▶ 2通り試した

- Wordform estimates ... 単語そのものをそのままカウントする
  - Data sparsity が問題
- Cluster estimates ... 単語をクラスタリング
  - Common Crawl で学習した GloVe を K-means でクラスタリング
  - 各クラスタが単一単語とならないようにして sparsity を解決
  - 意味的に近い単語をまとめる (blue と red は同じものとみなす)

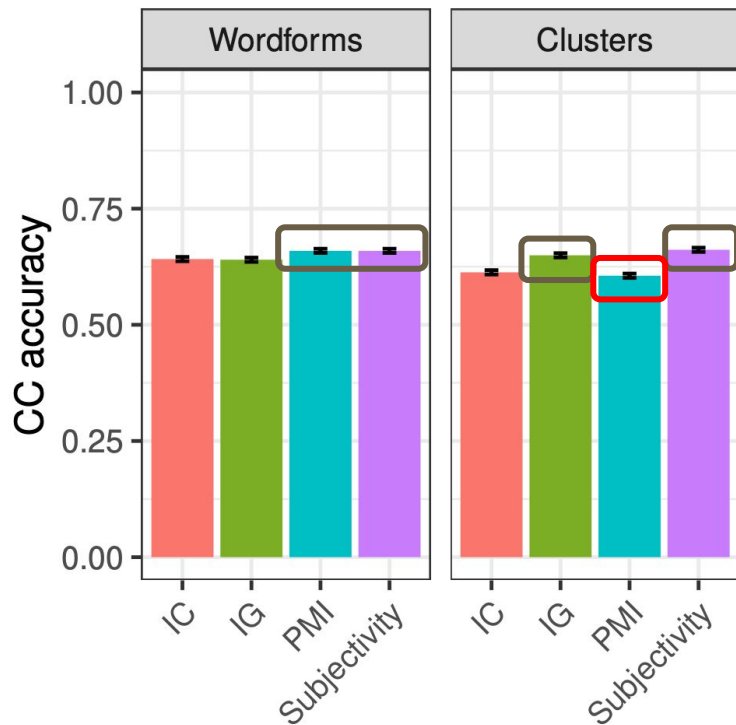
# Evaluation

- ▶ Predictor そのものを使うのではなく、Predictor を用いて分類器を作る
  - Logistic Regression ... スコア差を用いて回帰 (Morgan and Levy, 2016)

$$\log \frac{p(\langle A^1, A^2, N \rangle)}{p(\langle A^2, A^1, N \rangle)} = \beta_0 + \beta_1 (S(A^1, N) - S(A^2, N)) + \epsilon$$

- 複数 predictor を組み込んだ評価が可能となる
  - $\beta_2, \beta_3, \dots$ と増やしていけば良い

# Results ... 単一 predictor による精度評価



- Wordforms: subjectivity, PMI
- Clusters: Subjectivity, Information gain
- PMI predictor の精度は Cluster-based では落ちる
- IG は Cluster-based では上がる

(a) Common Crawl ( $N = 41822$ ).

# Results ... 複数 predictor による精度評価

Predictor	Accuracy	Conf. Interval
Subj. (cluster) + PMI (wordform)	<b>.723</b>	[.719, .727]
Subj. (wordform) + PMI (wordform)	.713	[.708, .717]
Subj. (cluster) + IG (cluster)	.699	[.695, .703]
Subj. (cluster) + IC (cluster)	.690	[.686, .695]
IG (cluster) + IC (cluster)	.684	[.680, .689]

- 2つの predictor の組み合わせを全通り試した(3つ以上は効果なし)
  - やはり Subj. と PMI が強い印象

# Results ... Quantitative analysis

## ▶ Collocation が含まれるようなケースの予測

例) "major bad behavior"

- PMI は他の predictor に比べて collocation をうまく扱う
- 逆に、collocation 的な NP が他の形容詞によって阻害されている場合、PMI は間違えた予測をしてしまう

例) "dear sick friend"

## ▶ cluster-based で PMI の精度が落ちる ⇐ collocation は語彙に依るため！

# 感想

- 個人的に流れが好み
  - なんらかの言語現象
  - それを説明しようとする理論がある
  - 計算機で比較・検証する
- 結果の解釈が難しい
  - Integration Cost に関しては精度低めだが、その辺りの議論が無かった
- 言語・文化圏によっても差がある？
  - typological な違い
  - 主観性 ...「青さ」の主観度が高い文化圏もあるかも