

2023年3月17日

言語処理学会ワークショップ

Deep Computational Linguistics

イントロダクション

「言語処理と言語学の橋渡しの面白さと難しさ」



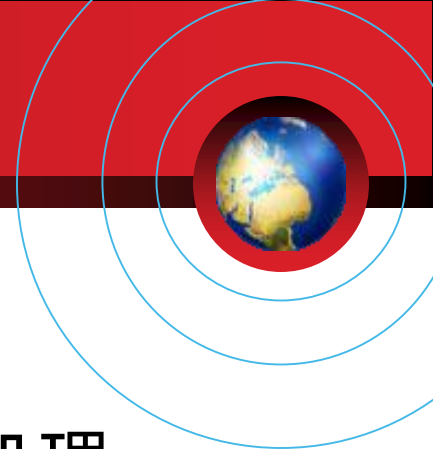
永田亮

甲南大学知能情報学部

はじめに  
少し自己紹介



# 少し自己紹介



## ■ 永田亮(甲南大学)



## ■ 専門

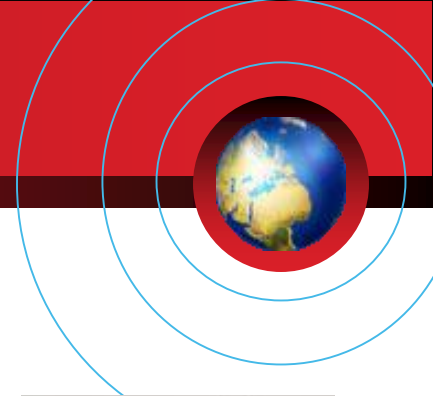
- 自然言語処理
- 計算言語学

## ■ 興味

- 語学学習
- 言語習得
- 言語のモデル化
- food&wine
- football

このworkshopを交流の場とできたらと思っています

# 言語系共同研究メンバ



永田  
(甲南大)  
語学学習支援  
言語習得



高村  
(産総研)  
言語生成  
自動要約



川崎  
(東大)  
西語歴史  
年代推定



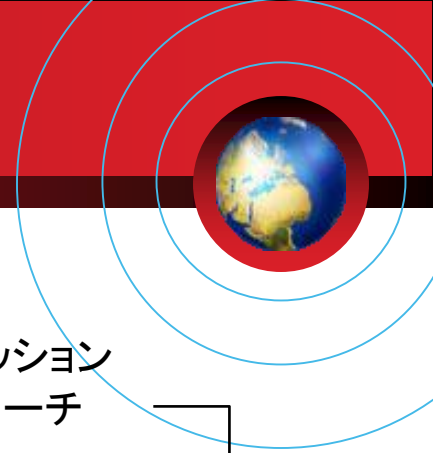
大谷  
(東京外大)  
談話機能言語学  
構文文法

個人&グループ経験: 言語処理の言語学への応用

timeline towards this workshop



# Before deep era



## World

2015: 言語処理学会テーマセッション  
言語探求のための数理的アプローチ  
with 高村+

2014: 言語処理学会テーマセッション  
言語探究の手段としての言語処理  
with 高村+



2015

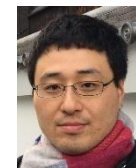
2010

2005

2000



高村氏と出会う  
(二人とも言語好き)



川崎氏と出会う  
(テーマセッションで発表)

2006: 英語名詞の  
可算性モデル化 (Nagata+)

2014: 非母語話者英語データ  
から印欧語の言語系統樹が再  
構築できることを発見 (Nagata+)

B: 電気工学(向いていない)  
仮想言語学  
(5人称, 現未来形...)

## Personal

# Advent of deep era



## World

2014: 言語処理学会テーマセッション  
言語探究の手段としての言語処理

2015: 言語処理学会テーマセッション  
言語探求のための数理的アプローチ

2015

2022:

統数研シンポジウム「現代語の意味の変化に対する  
計算的・統計力学的アプローチ」(持橋+)

英文学会シンポジア

自然言語処理と機械学習を用いた実証的な  
認知言語学の研究の可能性を探る

with 大谷, 高村, 川崎+

2023-: more to come!

We're here!

2020



2019: 講演@外語大  
大谷氏に会う

2017: 外来語の意味変化に対す  
る数理的分析 (高村, 永田, 川崎)

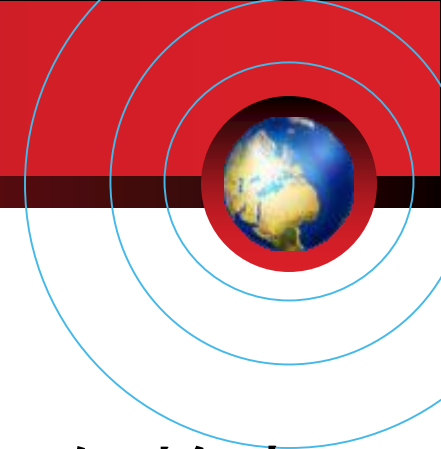
2016: 非母語話者英語専用の  
構文アノテーション方法 (Nagata+)

2022:

言語処理的アプローチによる  
better-off構文定着過程の説明  
(永田, 大谷, 高村, 川崎)

2023

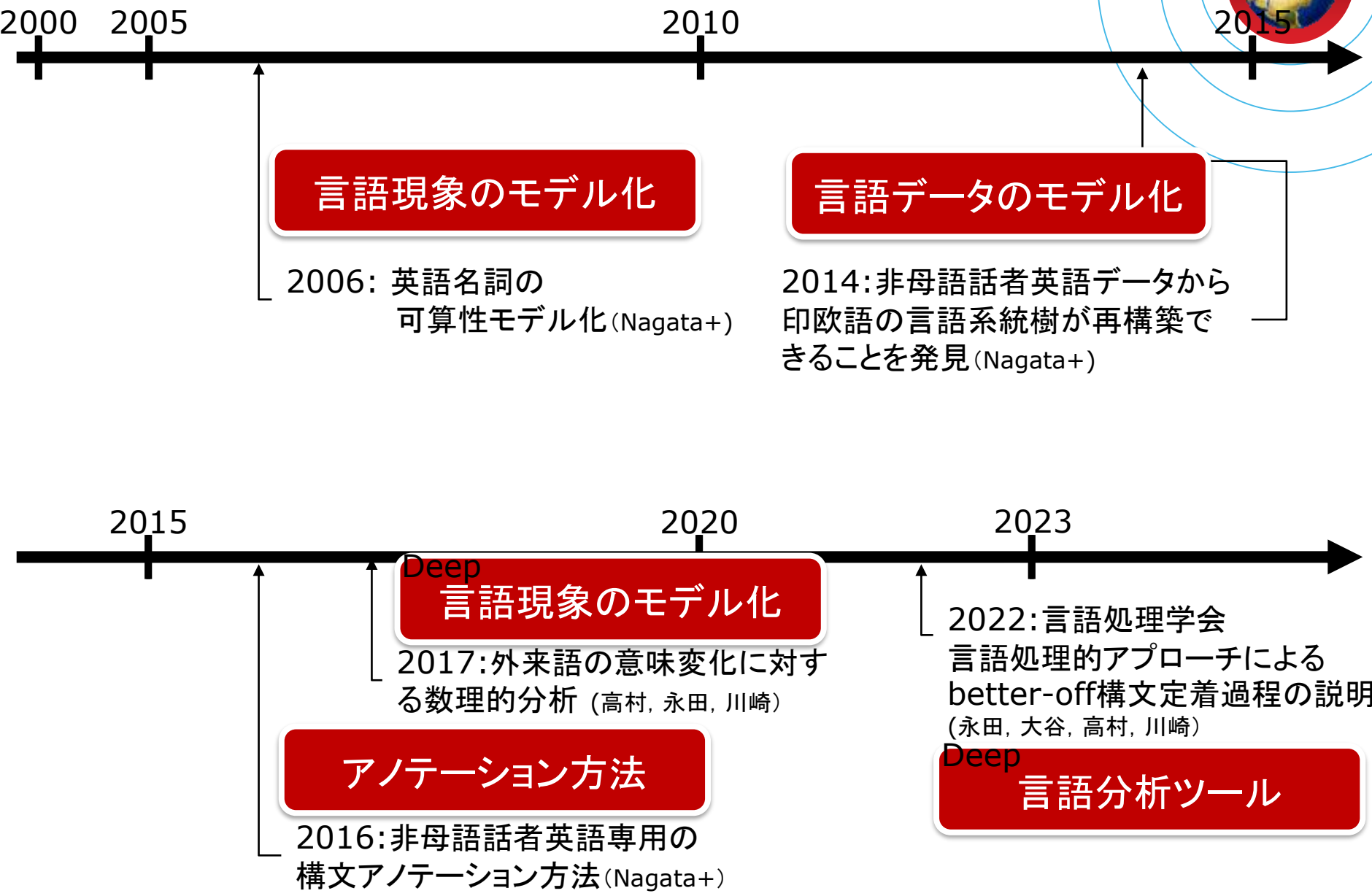
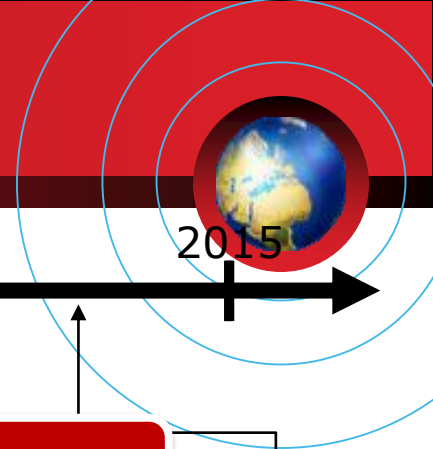
## Personal



- **言語分析のためのツール**
  - 言語情報抽出(例:品詞, 構文), 用例検索
- **言語現象の記述方法(アノテーション)**
  - 特に, 計算機で扱いやすい形式
- **言語現象／データのモデル化(説明)**
  - 各種予測
- **(人の)言語能力のモデル化**
  - 深層学習(特にベクトル／分散表現)に期待



# これまでの研究の分類



# 具体例

言語処理学会 第28回年次大会 発表論文集 (2022年3月)

---

## 言語処理的アプローチによる better off 構文の定着過程の説明

永田亮<sup>1</sup> 大谷直輝<sup>2</sup> 高村大也<sup>3</sup> 川崎義史<sup>4</sup>

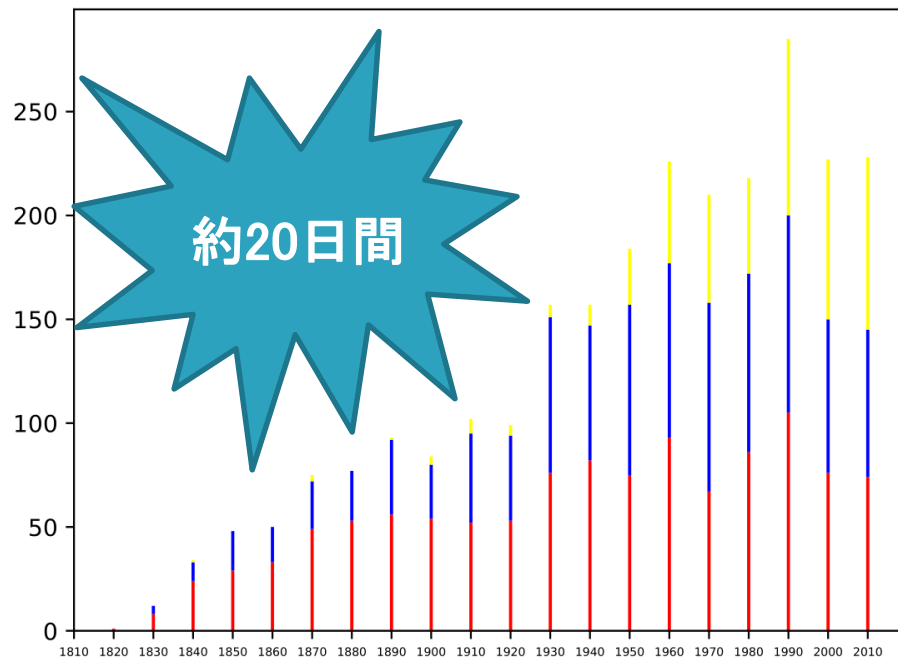
<sup>1</sup> 甲南大学 <sup>2</sup> 東京外国語大学 <sup>3</sup> 産業技術総合研究所 <sup>4</sup> 東京大学

nagata-better\_off @ hyogo-u.ac.jp.

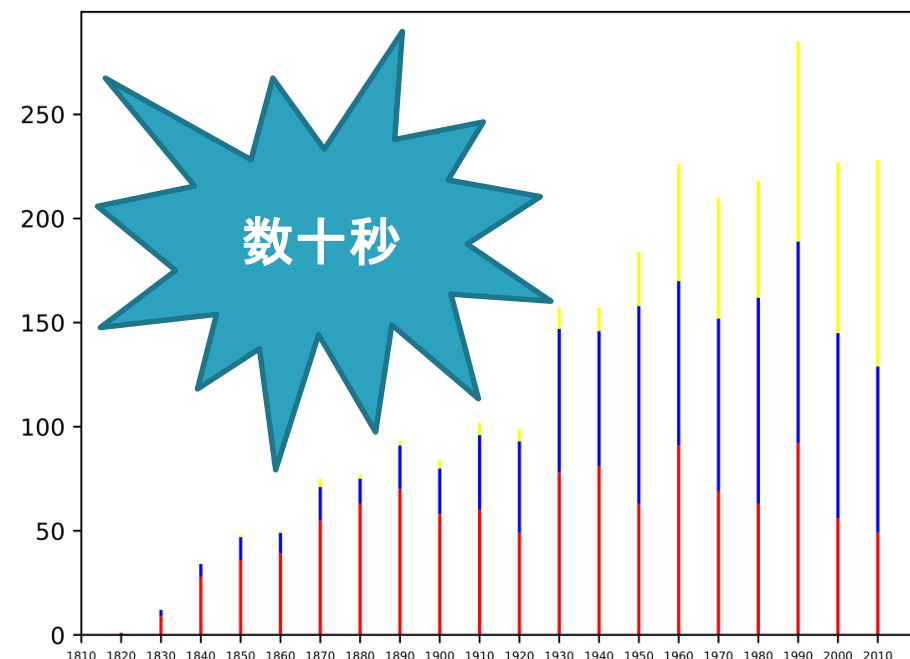
# 具体例：フレーズ better offの意味変化



人手で分析  
グルーピング



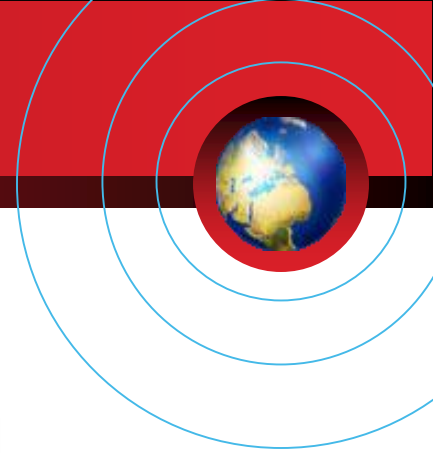
BERTでベクトル化  
自動グルーピング



客観的グルーピングが人手のグルーピングに一致

仮説の生成, サポートに活用できそう

# 個人的な面白さ, 期待

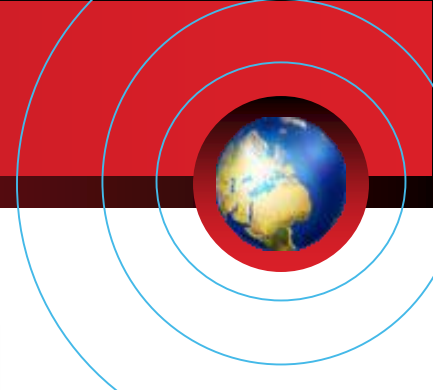


永田

(たまたまかもしれないけど),  
完全にデータドリブンに分類した  
結果が言語学者の分析に一致する  
のは興味深い

他にも色々使えそう:  
新語義発見, 意味変化の分析, 文法  
化の定量化, construalの定量化...

# 思い・感覚の違い



永田

熟練の言語学者の分析に  
勝るものはない(という思い)

言語分析をする上で個人の感覚のみ  
に頼ってもよいのか  
(客観性に欠けるという思い)

言語学者の直観とは独立して、デー  
タドリブンに何かが示せるのが嬉しい  
(用法基盤モデルとの相性◎)



大谷

# 期待度の違い



永田

深層学習／NLPはものすごくパワフル(すごいことが出来そう)

LMを、人間の知識構造／情報処理過程の近似値に見立てられないか

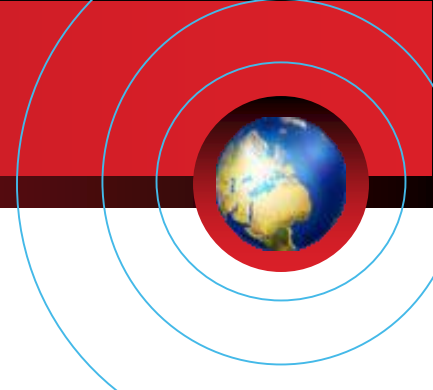
中で何が起きているのかを解析するのは非常に困難

現状で何かを証明するのはほぼ不可能(サポートにはなるかも)



大谷

# 期待度の違い(言い換えると)



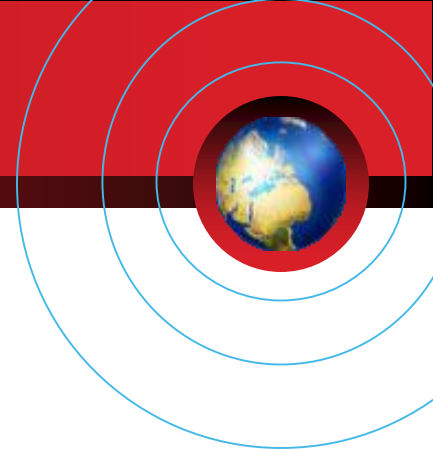
永田

深層学習／NLPのことはなんでも  
知っているだろう

言語学のことはなんでも  
知っているだろう



大谷



## ■ 本ワークショップに期待

- 様々な発表
- 最後の交流会（雑談会）

## ■ 今後のイベント

- 2023秋：某英語系学会のシンポジウム  
深層学習時代と英語研究に関するもの
- 2024：言語処理学会30周年記念事業  
言語処理技術セミナー special edition  
「人文系研究者のための言語処理」（仮）