



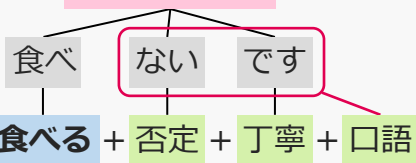
概要：日本語版UniMorphを作成

結果：国際的なスキーマに準拠して、日本語の動詞における語形変化とその意味をまとめたデータセットを作成した
 貢献：日本語を含め、言語横断的に利用できる資源を構築
 今後：データセット拡張、日本語学習者の支援ツールの開発

背景：日本語動詞の語形変化は多様

形態素解析しても…

食べないです



☹️ 語尾と意味 ≠ 1対1

食べる + 否定 + 丁寧 = 食べません

食べる + 過去 + 可能 = ???

召し上がった - 尊敬 = ???

語形を意味の足し算で表現したい!

日本語の107動詞, 計12,533組の語形と意味を対応付けた

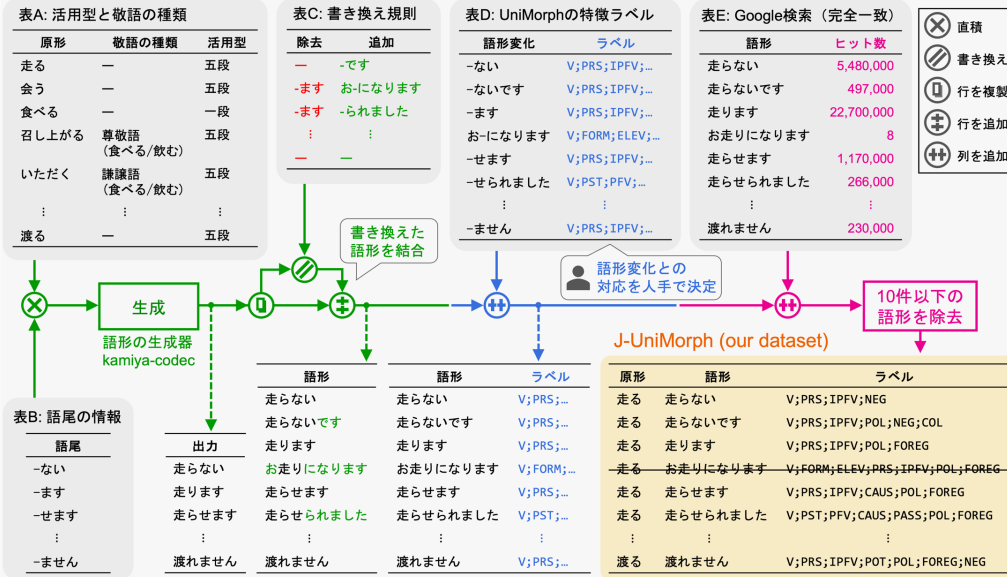
UniMorphとは?

- 各言語に共通のスキーマで動詞などの語形に特徴ラベルを付与
- 言語横断的な解析の改善を目指し, 169以上の言語で作成
- ☹️ 人手で作成した日本語のデータセットがなかった

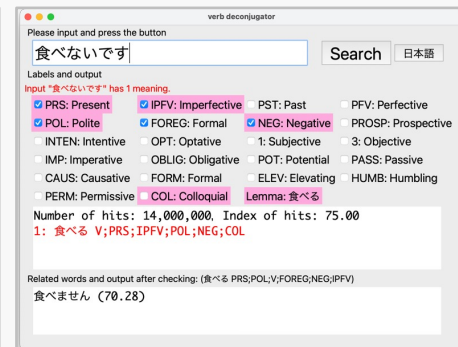
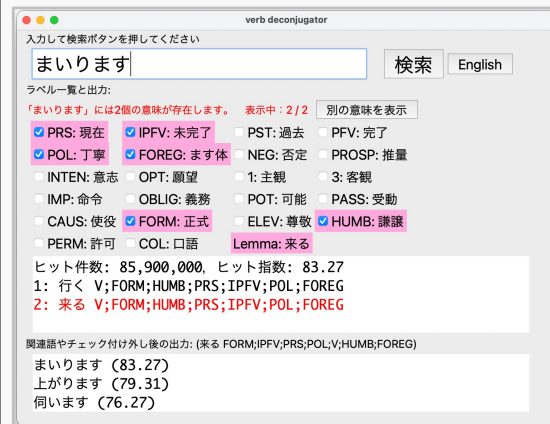
原形	語形	特徴ラベル	意味の足し算
書く	書きません	V; PRS; IPFV; POL; FOREG; NEG	
答える	答えられる	V; PRS; IPFV; POT	
食べる	お食べください	V; FORM; IMP; POL	
食べる	召し上がる	V; FORM; ELEV; PRS; IPFV	現在・未完了
来る	まいります	V; FORM; HUMB; PRS; IPFV; POL; FOREG	尊敬/謙讓, 丁寧

作成したUniMorphデータセットの一部

作成の流れ：語形の生成→ラベルの決定→フィルタリング



応用：日本語学習者向けツールの試作



$$\text{ヒット指数} = 100 \times \frac{\log_{10}(\text{hits} + 1)}{\log_{10}(\text{max_hits} + 1)}$$