

自然言語処理を用いた入力した文章の イントネーションの可視化



総合情報学科 知能情報システム研究室
永井ゼミ 4年 江澤 直樹



目次

- ➡ 動機と目的
- ➡ 利用環境
- ➡ アクセントの概要
- ➡ 東京式アクセントの種類
- ➡ 手順
- ➡ イメージ
- ➡ 研究計画
- ➡ 参考文献



動機と目的

動機

日本語を学ぶ外国人が文章や単語のアクセントを調べるものがあれば学習が楽になるのではないかと思います。

目的

入力された文章からイントネーションを出力するプログラムの作成



利用環境

➡ Python

➡ MeCab(ipadic)

アクセントの概要

- ・アクセント：単語ごとの音の高低
- ・イントネーション：文章全体の音の高低

- ・東京式アクセントの特徴

1. 第1拍と第2拍では必ず高さが変わる(拍：リズム上の単位のこと)

— — — — —
— — — — —
緑 ミドリ 平仮名 ヒラガナ

2. 1語の中で音の高さが急に下がるのは1回だけ

音の高さが急に下がる直前の音を「**アクセント核**」と呼ぶ

東京式アクセントの種類

- ・ 平板型(ヒライタ型)

— ()

桜 サクラ(が)

- ・ 起伏型

頭高型(アタマダカ型)

— ()

命 イノチ(が)

中高型(ナカダカ型)

— ()

心 ココロ(が)

尾高型(オダカ型)

— ()

雫 シズク(が)



手順

1. 文章を入力する
2. 入力した文章にMeCabを用いて形態素解析を行い、形態素、品詞、発音などに分解しする
3. 形態素解析を行った結果をリストの中へ入れる
4. 事前に準備しておいたアクセントのデータを読み込む。
5. 既にリスト化した形態素の単語や発音などと一致しているアクセントのデータを正規表現を用いて検索し、各形態素ごとのアクセントのデータを取り出す。
6. 単語同士が結合する際の規則に則って各単語のアクセントを変化させる
7. 入力した文章と、文章の発音、文章のイントネーションを出力する

イメージ①

卒業研究は大変だ。

形態素解析

卒業	名詞	ソツギョウ
研究	名詞	ケンキュウ
は	助詞	ハ
大変	名詞	タイヘン
だ	助動詞	ダ
。	記号	。

アクセントデータを割り振る

■ アクセントデータの例

卒業,名詞,ソツギョウ,LHHHH,尾高型

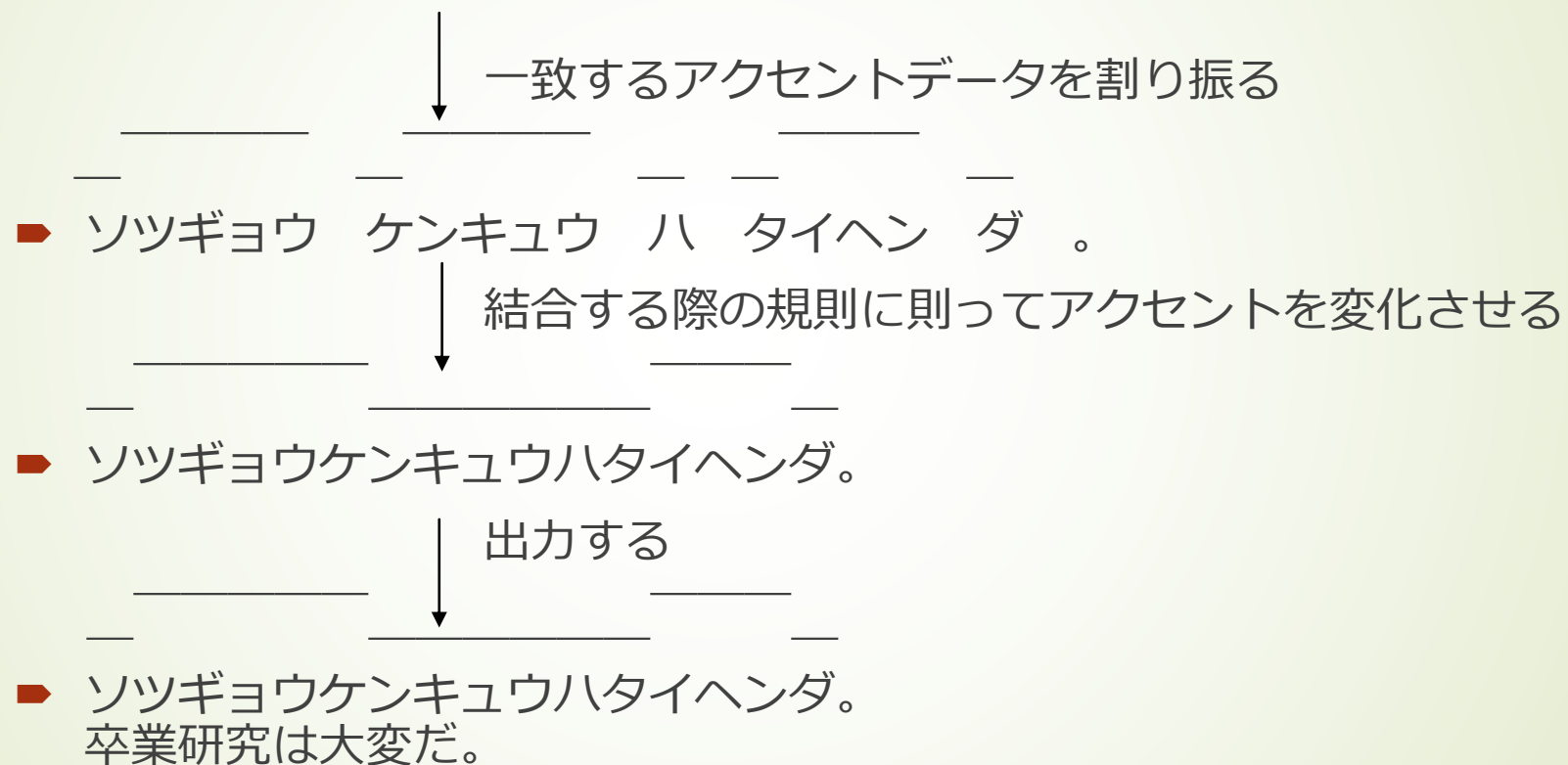
研究,名詞,ソツギョウ,LHHHH,尾高型

は,助詞,ハ,L

大変,名詞,タイヘン,LHHH,尾高型

だ,助動詞,ダ,L

イメージ②



研究計画

1	2	3	4	5	6	7	8	9	10	11	12	1
---	---	---	---	---	---	---	---	---	----	----	----	---

プログラミング学習

構想・設計

プログラム作成

アクセントデータ作成

検証

論文作成

参考文献

- ▶ **東京式アクセントについて(平板型・起伏型(頭高型・尾高型・中高型))**
<https://www.nihongo-appliedlinguistics.net/wp/archives/4519>
- ▶ **語のつながりのアクセント規則を用いた仮名文字文の形態素解析**
https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=129317&item_no=1&page_id=13&block_id=8
- ▶ **辞結合におけるアクセント自動生成規則**
https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=126223&item_no=1&page_id=13&block_id=8
- ▶ **日本語複合語のアクセント付与規則**
https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/45656/1/BISC009_005.pdf